

# Abstract

The origin and evolution of stellar populations is one of the greatest challenges in modern astrophysics. It is known that the majority of the stars has its origin in stellar clusters (Carpenter 2000; Porras et al. 2003; Lada & Lada 2003). However, only less than one tenth of these clusters remains bounded after the first few hundred million years (Lada & Lada 2003). Ergo, the understanding of the origin and evolution of stars demands meticulous analyses of stellar clusters in these crucial ages.

The project Dynamical Analysis of Nearby Clusters (DANCe, Bouy et al. 2013), from which the present work is part of, provides the scientific framework for the analysis of Nearby Young Clusters (NYC) in the solar neighbourhood ( $\leq 500$  pc). The DANCe carefully designed observations of the well known Pleiades cluster provide the perfect case study for the development and testing of statistical tools aiming at the analysis of the early phases of cluster evolution.

The statistical tool developed here is a probabilistic intelligent system that performs Bayesian inference for the parameters governing the probability density functions (PDFs) of the cluster population (PDFCP). It has been benchmarked with the Pleiades photometric and astrometric data of the DANCe survey. As any Bayesian framework, it requires the setting up of priors. To avoid the subjectivity of these, the intelligent system establish them using the Bayesian Hierarchical Model (BHM) approach. In it, the parameters of prior distributions, which are also inferred from the data, are drawn from other distributions in a hierarchical way.

In this BHM intelligent system, the true values of the PDFCP are specified by stochastic and deterministic relations representing the state of knowledge of the NYC. To perform the parametric inference, the likelihood of the data, given these true values, accounts for the properties of the data set, especially its heteroscedasticity and missing value objects. By properly accounting for these properties, the intelligent system: i) Increases the size of the data set, with respect to previous studies working exclusively on fully observed objects, and ii) Avoids biases associated to fully observed data sets, and restrictions to low-uncertainty objects ( $\sigma$ -clipping procedures).

The BHM returns the posterior PDFs of the parameters in the PDFCPs, particularly of the spatial, proper motions and luminosity distributions. In the BHM each object in the data set contributes to the PDFs of the parameters proportionally to its likelihood. Thus, the PDFCPs are free of biases resulting from typical high membership probability selections (sampling bias).

As a by-product, the BHM also gives the PDFs of the cluster membership probability for each object in the data set. These PDFs together with an optimal probability classification threshold, which is obtained from synthetic data sets, allow the classification of objects into cluster and field populations. This by-product classifier shows excellent results when

applied on synthetic data sets (with an area under the ROC curve of 0.99). From the analysis of synthetic data sets, the expected value of the contamination rate for the PDFCPs is  $5.8 \pm 0.2\%$ .

The following are the most important astrophysical results of the BHM applied to the Pleiades cluster. First, used as a classifier, it finds  $\sim 200$  new candidate members, representing 10% new discoveries. Nevertheless, it shows outstanding agreement (99.6% of the  $10^5$  objects in the data set) with previous results from the literature. Second, the derived present day system mass distribution (PDSMD) is in general agreement with the previous results of [Bouy et al. \(2015\)](#).

Thus, by better modelling the data set and eliminating unnecessary restrictions to it, the new intelligent system, developed and tested in the present work, represents the state of the art for the statistical analysis of NYC populations.

## French abstract

Il semble maintenant établi que la majorité des étoiles se forment dans des amas ([Carpenter 2000](#); [Porras et al. 2003](#); [Lada & Lada 2003](#)). Comprendre l'origine et l'évolution des populations stellaires est donc l'un des plus grands défis de l'astrophysique moderne. Malheureusement, moins d'un dixième de ces amas restent gravitationnellement liés au delà de quelques centaines de millions d'années ([Lada & Lada 2003](#)). L'étude des amas stellaires doit donc se faire avant leur dissolution dans la galaxie.

Le projet Dynamical Analysis of Nearby Clusters (DANCe, [Bouy et al. 2013](#)), dont le travail fait partie, fournit le cadre scientifique pour l'analyse des amas proches et jeunes (NYC) dans le voisinage solaire. Les observations de l'amas ouvert des Pléiades par le projet DANCe offrent une opportunité parfaite pour le développement d'outils statistiques visant à analyser les premières phases de l'évolution des amas.

L'outil statistique développé ici est un système intelligent probabiliste qui effectue une inférence bayésienne des paramètres régissant les fonctions de densité de probabilité (PDF) de la population de l'amas (PDFCP). Il a été testé avec les données photométriques et astrométriques des Pléiades du relevé DANCe. Pour éviter la subjectivité de ces choix des priors, le système intelligent les établit en utilisant l'approche hiérarchique bayésienne (BHM). Dans ce cas, les paramètres de ces distributions, qui sont également déduits des données, proviennent d'autres distributions de manière hiérarchique.

Dans ce système intelligent BHM, les vraies valeurs du PDFCP sont spécifiées par des relations stochastiques et déterministes représentatives de notre connaissance des paramètres physiques de l'amas. Pour effectuer l'inférence paramétrique, la vraisemblance (compte tenu de ces valeurs réelles), tient en compte des propriétés de l'ensemble de données, en particulier son hétérosécédasticité et des objects avec des valeurs manquantes.

Le BHM obtient les PDF postérieures des paramètres dans les PDFCP, en particulier celles des distributions spatiales, de mouvements propres et de luminosité, qui sont les objectifs scientifiques finaux du projet DANCe. Dans le BHM, chaque étoile du catalogue contribue aux PDF des paramètres de l'amas proportionnellement à sa probabilité d'appartenance. Ainsi, les PDFCP sont exempts de biais d'échantillonnage résultant de sélections tronquées au-dessus d'un seuil de probabilité défini plus ou moins arbitrairement.

Comme produit additionnel, le BHM fournit également les PDF de la probabilité d'appartenance à l'amas pour chaque étoile du catalogue d'entrée, qui permettent d'identifier les membres probables de l'amas, et les contaminants probables du champ. La méthode a été testée avec succès sur des ensembles de données synthétiques (avec une aire sous la courbe ROC de 0,99), ce qui a permis d'estimer un taux de contamination pour les PDFCP de seulement 5,8 %.

Ces nouvelles méthodes permettent d'obtenir et/ou de confirmer des résultats importants sur les propriétés astrophysiques de l'amas des Pléiades. Tout d'abord, le BHM a

découvert 200 nouveaux candidats membres, qui représentent 10% de la population totale de l'amas. Les résultats sont en excellent accord (99,6% des 100 000 objets dans l'ensemble de données) avec les résultats précédents trouvés dans la littérature, ce qui fournit une validation externe importante de la méthode. Enfin, la distribution de masse des systèmes actuelle (PDSMD) est en général en bon accord avec les résultats précédents de [Bouy et al. \(2015\)](#), mais présente l'avantage inestimable d'avoir des incertitudes beaucoup plus robustes que celles des méthodes précédentes.

Ainsi, en améliorant la modélisation de l'ensemble de données et en éliminant les restrictions inutiles ou les hypothèses simplificatrices, le nouveau système intelligent, développé et testé dans le présent travail, représente l'état de l'art pour l'analyse statistique des populations de NYC.

# Contents

Abstract . . . . .	i
French abstract . . . . .	iii
<b>1 Introduction</b>	<b>1</b>
1.1 The initial mass function of stellar clusters . . . . .	2
1.2 Numerical simulations of the early stages of star formation . . . . .	7
1.3 The DANCe project . . . . .	8
1.4 Methodologies for the classification of cluster members . . . . .	11
1.5 A new intelligent system: the Bayesian Hierarchical Model . . . . .	16
<b>2 The Pleiades as a benchmark</b>	<b>20</b>
2.1 Generalities . . . . .	20
2.2 The distance to the Pleiades . . . . .	20
2.3 Projected spatial distribution . . . . .	22
2.4 Velocity Distribution . . . . .	26
2.5 Luminosity Distribution . . . . .	28
2.6 Mass Distribution . . . . .	32
2.7 The Pleiades DANCe DR2 . . . . .	36
<b>3 Bayesian formalism</b>	<b>45</b>
3.1 Introduction to probability theory. . . . .	45
3.2 Bayesian Hierarchical Models . . . . .	54
3.3 The BHM for the Pleiades DANCe data set . . . . .	57
3.4 The methodology for the analysis of the PSD . . . . .	72
3.5 Priors . . . . .	79
3.6 Sampling the posterior distribution . . . . .	83
3.7 Codes . . . . .	92
<b>4 Results</b>	<b>100</b>
4.1 Performance of the classifier . . . . .	100
4.2 Comparison with the literature . . . . .	110
4.3 The statistical distributions of the Pleiades cluster. . . . .	117
4.4 Updating the prior knowledge . . . . .	121

4.5	Projected spatial distribution . . . . .	122
4.6	Proper motions distribution . . . . .	130
4.7	Luminosity distribution . . . . .	132
4.8	Mass distribution . . . . .	138
4.9	The mass distribution on time . . . . .	144
<b>5</b>	<b>Conclusions and future work</b>	<b>150</b>
5.1	Current issues and future improvements . . . . .	152

# Chapter 1

## Introduction

The majority of the stars are born together. Carpenter (2000) reports that 50 – 70% of very young,  $\leq 10$  millions of years (Myr), and 25 – 70% of the young,  $\leq 100$  Myr, stellar populations are formed in groups. Porras et al. (2003) and Lada & Lada (2003) find that among 80% to 90% of the stars are formed in groups with more than 100 members, which are called star clusters. Furthermore, as indicated by the former authors, these populous clusters represent 22% of the regions where stars form. The remaining of the star forming regions are small associations, with 5 to 30 members, where only up to 10% of the stars are formed. However, only less than 7% of these populous clusters survives as gravitationally bounded clusters when reaching an age of a few hundred Myr (Lada & Lada 2003). The remaining 93% of the star forming regions will become unbound and their stars will freely populate the galaxy. Thus, to understand the general rules that govern how the majority of stars forms, as well as the properties of the stars that populate our galaxy, it is crucial to fully decode the formation and early evolution of stellar clusters.

Astrophysicists, like archeologists and palaeontologists, can not willingly reproduce the vast majority of their studied phenomena. Although some experiments can be performed in specific situations (e.g. the chemical and physical properties of dust and gas), astrophysics remains an observational science. For this reason, to test the validity of their hypotheses, astrophysicists rely on statistical studies carried out over carefully designed observations. In particular, the understanding of the star formation process requires carefully designed observations of stellar clusters whose ages cover the early stages of their evolution.

The objective of this work is the construction, test and validation of a statistical tool, an intelligent system specifically, that given the carefully designed data of a stellar cluster, recovers the statistical distributions of its populations. In particular, it should deliver the cluster luminosity distribution, which can be transformed into the mass distribution given an evolutionary model and the cluster age. The mass distribution is a fundamental product of the star formation process. It contains the fingerprints of the early phases of star formation and subsequent cluster evolution.

An homogeneous and precise mass distribution inventory for clusters of diverse ages

and forming environments will allow the astrophysical community to test the current theories of the star formation process. In particular, it will allow to solve the questions about universality of the initial mass function (IMF, see next Section) and the role played by the physical properties of the cluster environment.

The remaining of this Chapter is structured as follows. In Section 1.1, I describe the importance of the initial mass distribution and some of its current models. In Section 1.2, I report the current status of numerical simulations of cluster formation and its impact on the understanding of the star formation process. In Section 1.3, I describe the project DANCe, its objectives and its carefully designed observations of stellar clusters. This work is part of that project and makes use of its observations. In Section 1.4, I comment on the past and current methodologies to select members of clusters and associations. Finally in Section 1.5, I briefly describe the methodology adopted for this new statistical tool, and its advantages over the previous works.

## 1.1 The initial mass function of stellar clusters

In his seminal work (Salpeter 1955), Edwin Salpeter defined the *original mass function*,  $\xi(M)$ , as

$$dN = \xi(M)d(\log_{10} M)\frac{dt}{T_0},$$

where  $dN$  is the number of stars in the mass range  $dM$  created in the time interval  $dt$  per cubic parsec, and  $T_0$  is the age of the galaxy. Following Chabrier (2003a), the mass function (MF) at the observed time  $t$ , is

$$\xi(\log_{10} M) = \frac{dN}{\log_{10} M},$$

where  $N$  is the stellar number density, and  $M$  the mass. The Initial Mass Function (IMF) is defined as the MF at the time of stellar formation  $t = t_0$ . The logarithmic transformation of the mass,

$$\xi(M) = \frac{1}{M \ln 10} \xi(\log_{10} M),$$

is convenient due to the large range of masses covered by the star formation process.

Notice that neither the IMF nor the MF are probability density functions (PDFs) of the mass (see Section 3.1 for the definition of a PDF). Nevertheless, they can be transformed into PDFs by a normalisation constant, which can be computed by integrating them as functions of the mass over the mass domain. In this work, I will use the PDF of the mass (PDM) in the logarithmic scale, denoted by  $\xi_L(\log_{10} M)$ , as a proxy for the MF. Thus,

$$\xi_L(\log_{10} M) \propto \xi(\log_{10} M).$$

The measuring and understanding of the IMF is a central topic in the study of star formation. It is also essential in other areas of astrophysics, from planetary formation, where it appears that the mass of the host star plays an important role in the formation of the planetary system (see for example [Mulders et al. 2015](#)), to galactic evolution ([Kennicutt 1998](#)) and cosmology (see for example [Narayanan & Davé 2012](#)).

The theories that predict the origin of the IMF can be categorised into deterministic and stochastic ([Offner et al. 2014](#)). The former postulate that stellar masses are deterministically inherited from the initial core masses via accretion from the gas reservoir of the parent molecular cloud. Thus, the IMF can be directly mapped from the distribution of initial core masses, and the understanding of the former reduces to that of the latter. On the other hand, stochastic models postulate that the stellar masses are independent of the initial core masses. Among these models, there are those proposing that stellar masses are determined by dynamical interactions and competitive accretion. For more details see [Offner et al. \(2014\)](#) and references therein.

The observational studies of the IMF are conditioned on the ages of the stellar populations under analysis (their MF at their corresponding ages), and rely deeply on the assumed processes that link the observed present-day MF to the IMF. While the resulting models for the IMF are analytical functions of the mass, the observed MF are commonly expressed with points, histograms or kernel density estimations<sup>1</sup> (KDEs).

The most common forms to describe the IMF are the power-law functions of Salpeter ([Salpeter 1955](#)), Miller and Scalo ([Miller & Scalo 1979](#)), and Kroupa ([Kroupa 2001, 2002; Kroupa et al. 2013; Thies & Kroupa 2007, 2008](#)), and the log-normal functions of Chabrier ([Chabrier 2003b,a, 2005](#)). Other functional forms include the truncated exponential ([de Marchi & Paresce 2001](#)), the Pareto-Levy family distribution ([Cartwright & Whitworth 2012](#)), and a log-normal with power-laws at high and low mass ranges ([Maschberger 2013](#)). For the sake of simplicity, I will only explain the classical ones of Salpeter, Chabrier, and Kroupa.

[Salpeter \(1955\)](#) derived his famous IMF using a luminosity function resulting from the compilation of the works of [Luyten \(1939, 1941\)](#) and [van Rhijn \(1925, 1936\)](#). Then, he transformed it into a MF using a mass-luminosity relation that he obtained after adopting a series of masses and luminosities from the literature. His MF has the form

$$\xi(M) = 0.03 \left( \frac{M}{M_\odot} \right)^{-1.35},$$

with  $M$  in the range  $0.3 M_\odot$  to  $17 M_\odot$ . Its units are  $M_\odot^{-1} \cdot pc^{-3}$ .

---

<sup>1</sup>KDEs are non-parametric ways to estimate a probability density function by means of an independently and identically distributed sample drawn from it.

[Chabrier \(2003b,a\)](#) derived his Present-day MF (PDMF) from the nearby luminosity functions of both the  $V$  band ([Dahn et al. 1986](#)) and the  $K$  band ([Henry & McCarthy 1990](#)) for the solar neighbourhood ( $\leq 5$  pc). He used the [Delfosse et al. \(2000\)](#) and [Baraffe et al. \(1998\)](#) mass-magnitude relations in  $V$  and  $K$  bands, respectively, to transform the luminosity functions into masses. Then, he fitted a log-normal form to the single objects with masses below  $1 M_\odot$ . The PDMF he found, in units of  $(\log M_\odot)^{-1} \cdot pc^{-3}$ , is

$$\xi(\log m)_{m \leq 1M_\odot} = 0.158_{-0.046}^{+0.051} \times \exp \left\{ -\frac{(\log m - \log 0.079_{-0.016}^{+0.021})^2}{2 \times (0.69_{-0.01}^{+0.05})^2} \right\}.$$

As shown by [Scalo \(1986\)](#),  $1 M_\odot$  is the limit at which the PDMF starts to differ from the IMF. Therefore, [Chabrier \(2003a\)](#) uses his PDMF as the IMF below the  $1 M_\odot$  limit. Above it, he adopts the Salpeter IMF,

$$\xi(\log m)_{m > 1M_\odot} = 4.43 \times 10^{-2} \cdot m^{-1.3 \pm 0.3}.$$

As shown by [Kroupa et al. \(1991\)](#), the discrepancies between the luminosity functions derived from photographic samples and from trigonometric parallaxes of nearby stars can be accounted with unresolved binaries. For this reason, [Chabrier \(2003b\)](#) derived the Present-Day system mass function (PDSMF) for unresolved systems. It takes into account each object as a possible unresolved binary or system. This PDSMF has the form,

$$\xi(\log m)_{m \leq 1M_\odot} = 0.086 \times \exp \left\{ -\frac{(\log m - \log 0.22)^2}{2 \times (0.57)^2} \right\},$$

with the same normalisation and coefficients of the IMF above  $1 M_\odot$ .

Later, [Chabrier \(2005\)](#) included in his analysis the revision that [Reid et al. \(2002\)](#) made to the sample of ([Dahn et al. 1986](#)), and the extended sample of 8 pc from [Reid et al. \(2004\)](#). His new PDMF is

$$\begin{aligned} \xi(\log m)_{m \leq 1M_\odot} &= 0.093 \times \exp \left\{ -\frac{(\log m - \log 0.2)^2}{2 \times (0.55)^2} \right\}, \\ \xi(\log m)_{m > 1M_\odot} &= 0.041 \times m^{-1.3 \pm 0.3}. \end{aligned}$$

And his new PDSMF is

$$\begin{aligned} \xi(\log m)_{m \leq 1M_\odot} &= 0.076 \times \exp \left\{ -\frac{(\log m - \log 0.25)^2}{2 \times (0.55)^2} \right\}, \\ \xi(\log m)_{m > 1M_\odot} &= 0.041 \times m^{-1.3 \pm 0.3}. \end{aligned} \tag{1.1}$$

The canonical IMF of Kroupa ([Kroupa et al. 2013](#)) is a three-segment power-law, with two segments describing the IMF of the stars, while the third does it for Brown-Dwarfs

(BD)<sup>2</sup>. It has the following analytical representation

$$\xi_{BD}(m) = \frac{k}{3} \cdot \left( \frac{m}{0.07} \right)^{-0.3 \pm 0.4}, \quad 0.01M_\odot < m \leq 0.15M_\odot, \quad (1.2)$$

$$\begin{aligned} \xi(m) &= k \cdot \left( \frac{m}{0.07} \right)^{-1.3 \pm 0.3}, & 0.07M_\odot < m \leq 0.5M_\odot, \\ \xi(m) &= k \cdot \left( \frac{0.5}{0.07} \right)^{-1.3 \pm 0.3} \cdot \left( \frac{m}{0.5} \right)^{-2.3 \pm 0.36}, & 0.5M_\odot < m \leq 150M_\odot, \end{aligned} \quad (1.3)$$

where  $k$  is a constant.

[Thies & Kroupa \(2007\)](#) explored the parametric space of the IMF for the BD population. They performed a  $\chi^2$  fit of the maximum BD mass ( $m_{max,BD}$ ), the slope of the BD IMF ( $\alpha_{BD}$  in Eq. 1.2), and the population ratio ( $\mathcal{R}_{pop}$ ), which is the ratio of individual BD to individual stars. They performed this fit on literature data from the Trapezium, Taurus, IC 348 and the Pleiades nearby clusters. Figure 1.1, which reproduces their Figure 7, shows their results in these four clusters. In their work, the stellar IMF is the canonical one (Eq. 1.3). In the case of the Pleiades and IC 348, the slope of the BD IMF remained the canonical value  $\alpha_{BD} = 0.3$  due to the sparse data, and the Pleiades incomplete data particularly. The resulting system IMF is also shown as the solid curve. It results from the addition of the two IMFs (BD and stars) but without an overlap or discontinuity between them (see [Thies & Kroupa 2007](#), for details).

As shown in this Section, the IMF has been the subject of several studies which have transformed our knowledge about it. This knowledge evolved from the simple and efficient model of [Salpeter \(1955\)](#) in the high mass range, to the more complicated and still controversial models of [Thies & Kroupa \(2007\)](#); [Kroupa et al. \(2013\)](#) for the BD and low-mass stars. In the literature, the IMF of Chabrier ([Chabrier 2003b,a, 2005](#)) is often the preferred one. Although these models currently explain many of the observed phenomena, our knowledge of the physical processes that lead to the observed IMF is still incomplete. In an attempt to address the latter, the numerical simulations of the star formation process have done great progress. The next section is a tiny review of what these simulations have achieved in the past years.

---

<sup>2</sup>Brown-dwarfs are substellar objects with masses in the range from 10 to 80 Jupiter masses ( $0.01M_\odot$  to  $0.075 M_\odot$ ). Because there is no fusion of hydrogen in their cores, these objects are not classified as stars.

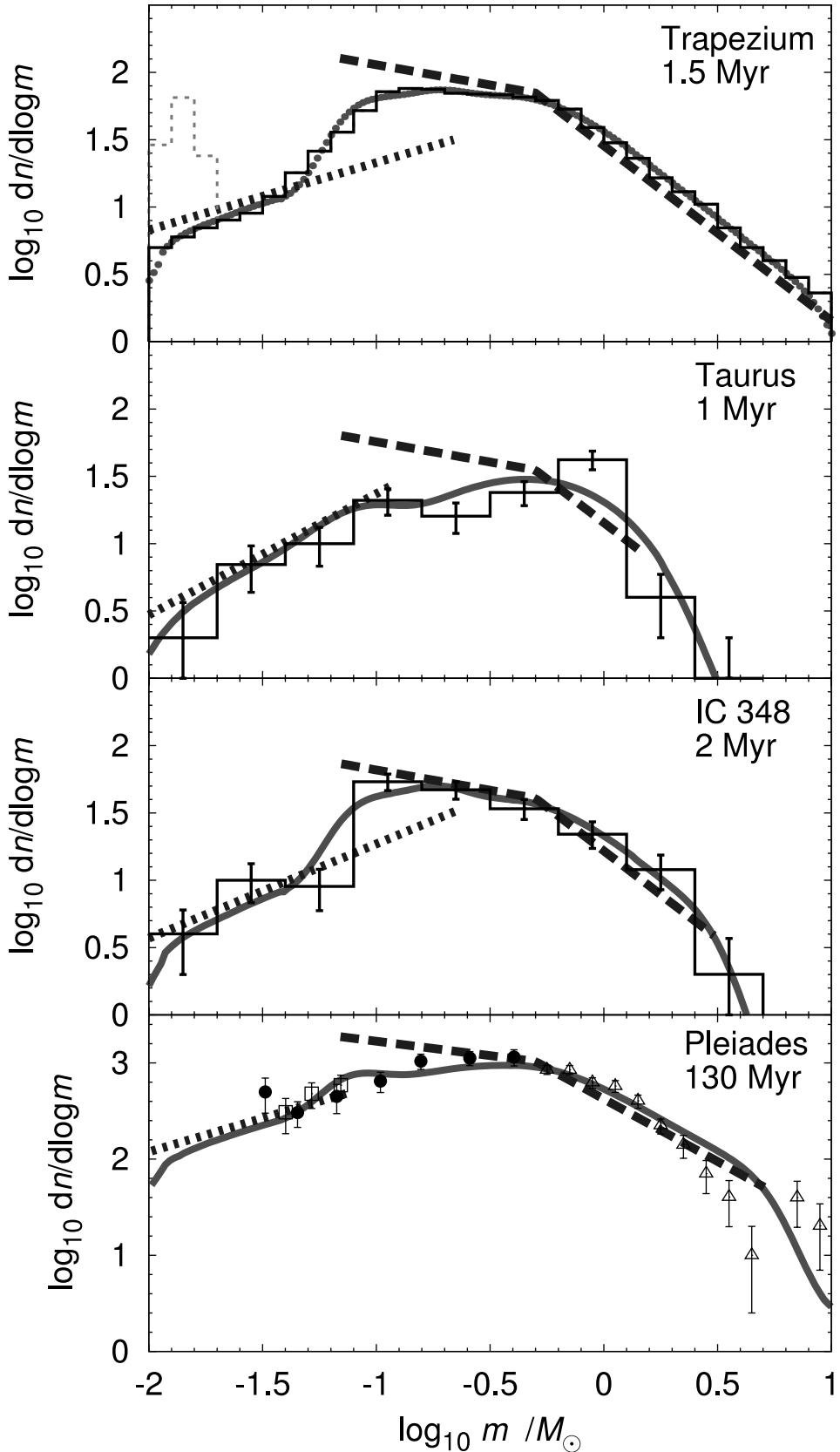


Figure 1.1: Observed mass distribution (histograms and dots) of the Trapezium, Taurus, IC348 and the Pleiades nearby clusters (from top to bottom). Also shown are the canonical stellar (long dashed line), fitted BD (dotted), and system (solid curve) IMFs of Thies & Kroupa (2007) . Reproduced from Figure 7 of Thies & Kroupa (2007), *A Discontinuity in the Low-Mass Initial Mass Function*, ApJ, Vol. 671

## 1.2 Numerical simulations of the early stages of star formation

In the first decade of this century, numerical simulations of star forming regions have proven to be of paramount importance in decoding the very early stages of the star formation process (e.g [Bate et al. 2003](#); [Jappsen et al. 2005](#); [Bate 2009a,c,b](#)). For example, [Bate et al. \(2003\)](#) using smooth particle hydrodynamics were able to simulate the collapse and fragmentation of a large-scale ( $50M_{\odot}$  within 0.375 pc radius) turbulent molecular cloud to form a stellar cluster. During the very first 0.1 Myr, which was the time covered by their simulation, they were able to simultaneously form discs and binary stars. The cloud formed roughly equal numbers of stars and BD (23 and 27, respectively) resulting in a mass distribution with a flat slope in the range  $0.01 - 0.5M_{\odot}$  (see Fig. 1.2). [Offner et al. \(2014\)](#) provides a review of the stellar initial mass distribution, and of the physical effects included in numerical simulations (radiative feedback, competitive accretion, dynamical interactions, magnetic fields) particularly.

In recent years, the works of [Kuznetsova et al. \(2015\)](#) and [Ballesteros-Paredes et al. \(2015\)](#), using the cold collapse paradigm (neglecting magnetic fields, radiative transfer and feedback), were able to probe that the main source driving the star formation process is gravity. Their simulations were typically run until 0.85 Myr in a box of 3 pc of side, and with masses in the few thousands of  $M_{\odot}$ . The mass distribution obtained by [Kuznetsova et al. \(2015\)](#) reproduces reasonably well the current models of the initial mass distribution. In particular, the IMF of [Chabrier \(2005\)](#) down to  $0.1 M_{\odot}$  (see Fig. 1.3). However, they produce too few low-mass stars and brown dwarfs.

Numerical simulations are of great use in the understanding of the star formation process, in the very early phases ( $\leq 1$  Myr) of its evolution particularly. Despite the fact that many of these simulations are in agreement with the observed mass distributions, currently they i) do not incorporate all astrophysical effects, ii) resolve close binaries and multiple systems, and iii) produce enough stellar objects to improve the statistics ([Offner et al. 2014](#)). Furthermore, these simulations require input from the observations (e.g. density, turbulence and magnetic fields in the molecular cloud). More importantly, they require feedback from the observations in order to fine tune the parameters of the physical processes. Real and simulated data must be compared in order to improve the latter. Thus, to constrain the current theories of the star formation process, precise and detailed studies of the early phases of cluster formations are still required. This is the objective of the DANCe project, from which the present work is part of.

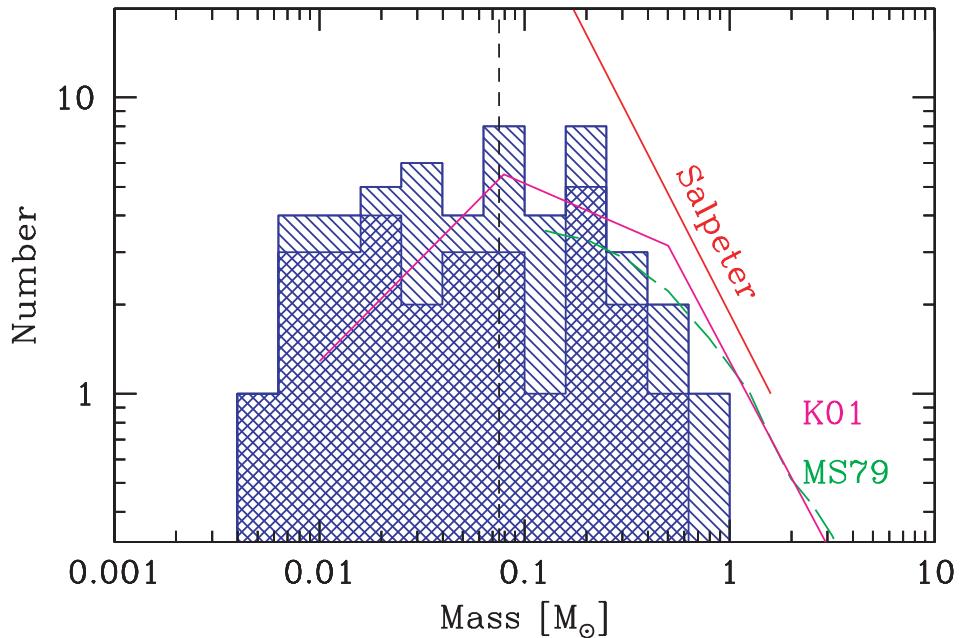


Figure 1.2: Mass distribution resulting from the numerical simulation of Bate et al. (2003). The lines show the mass distributions of Salpeter (1955), Miller & Scalo (1979) and Kroupa (2001). Reproduced from Figure 10 of Bate et al. (2003), *The formation of a star cluster: predicting the properties of stars and brown dwarfs*, MNRAS, Vol. 339

### 1.3 The DANCe project

The DANCe (Dynamical Analysis of Nearby Clusters<sup>3</sup>) project is an international and multidisciplinary collaboration of researchers whose main objective is *unravelling the origin of the mass function* of stellar Nearby Young Clusters (NYC). This objective demands a thorough compilation of resolved star forming populations over the entire mass spectrum and in diverse star forming environments. It requires complete catalogues of the star cluster members, and only those members. Disentangling star cluster members from the field population is a titanic task that the astrophysical community is on the verge of completing thanks to the future *Gaia* (Gaia Collaboration et al. 2016a) data. However, even with Gaia, the development and success of star formation theories will be pending a comprehensive identification of the sub-stellar populations and the internal dynamics of embedded cluster cores. *Gaia* will leave aside the most massive BD and planets, reaching only down to  $0.03 M_{\odot}$  on the best cases (Sarro et al. 2013). Furthermore, *Gaia* reaches the 20 mag limit in the *G* band (350 nm to 1000 nm Jordi et al. 2010) what makes it almost blind to the dust and gas in which the majority of the star forming regions are still embedded.

The DANCe photometric and astrometric measurements will complement *Gaia*'s exquisite astrometric precision in a selected list of star forming regions, which is shown in

---

<sup>3</sup><http://project-dance.com>

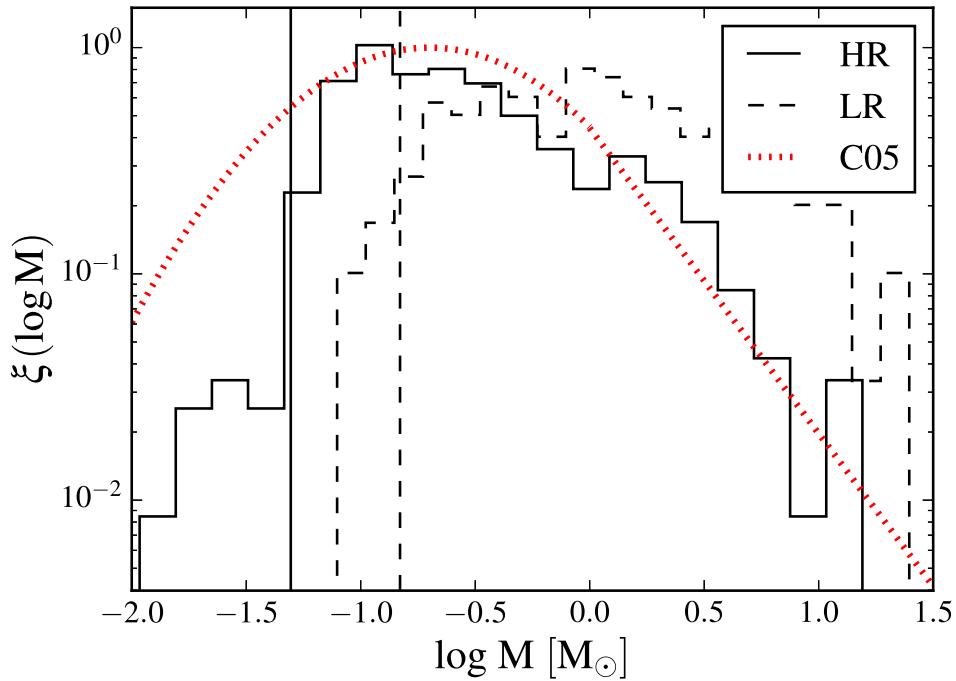


Figure 1.3: Mass distribution resulting from the numerical simulation of [Kuznetsova et al. \(2015\)](#). The High (solid line) and Low (dashed line) resolution simulations reach  $0.05 M_\odot$  (solid vertical line) and  $0.15 M_\odot$  (dashed vertical line), respectively. Also shown the normalised IMF of [Chabrier \(2005\)](#) (red dashed line). Reproduced from Figure 11 of [Kuznetsova et al. \(2015\)](#), *Signatures of Star Cluster Formation by Cold Collapse*, ApJ, Vol. 815

Table 1.1. In addition, its secondary targets are also listed in Table 1.2. A pilot study of the DANCe data properties has already been conducted on the Pleiades cluster. For more details on it I refer the reader to Section 2.7 and to [Bouy et al. \(2013\)](#).

The unravelling of the mass function requires in addition to accurate, precise, and complete data sets of NYC, a statistical methodology able to deal not just with the properties of the data sets and the particularities of each NYC, but with the capacity to disentangle the few hundreds cluster members from within the millions of sources of the field population. Working with individual stars one by one is no longer possible. Under the current paradigm of artificial intelligence, the intelligent systems offer a viable solution to the classical needle in a haystack problem. The DANCe project has already successfully tested its first intelligent classifier. Its methodology and astrophysical results on the Pleiades clusters are detailed in the works of [Sarro et al. \(2014\)](#) and [Bouy et al. \(2015\)](#), respectively.

With both the precise and accurate data and methodology, the DANCe project will be able to not just complement *Gaia* in the low-mass regime, but also deliver a complete inventory of the early phases of cluster formation in different environments. It will unravel the roles of the environment and initial condition in the final mass distribution, which is the ultimate goal of the DANCe project.

Table 1.1: Main targets

Name	R.A. (J2000)	Dec (J2000)	Distance (pc)	Age (Myr)	$\mu_\alpha \cos(\delta)$ (mas/yr)	$\mu_\delta$ (mas/yr)
Pleiades	56.750	+24.117	120	120	19	-44
IC2391	130.133	-53.033	175	50	-25	23
IC2602	160.742	-64.400	160	50	-22	10
$\gamma^2$ Vel	122.383	-47.336	350	5~10	-6	10
NGC 2547	122.6071	-49.1675	450	20~35	-6	10
IC 348	56.142	+32.163	~300	2	7	-9
NGC 1333	52.258	+31.348	~300	1	?	?
CrA	285.462	-36.982	130	1	?	?
$\eta$ -Cha	130.525	-79.027	97	5~9	-30	28
$\epsilon$ -Cha	179.90657	-78.22184	111	5~9	-38	-1
Upper Scorpius	243.0	-23.4	120	5~10	-9	-25
Lupus I	235.7587	-34.1517	140	1~3	~-15	~-25
Lupus II	239.283	-37.785	140	1~3	~-15	~-25
Lupus III	242.40	-39.05	140	1~3	~-15	~-25
Serpens	277.487	+01.240	~400	1~3	?	?
Ophiuchus	247.025	-24.542	125	1~3	-6	-25
Taurus	70.25	+25.87	140	1~3	8	-21
Blanco 1	1.029	-29.833	250	~100	21	3
Orion	83.8221	-05.3911	450	1~3	1.7	-0.3
NGC2451A	116.35	-37.97	197	50~80	-22	14
NGC2451B	116.35	-37.97	360	50~80	-10	14
CygOB2	308.30	+41.32	1450	1~7	-1.6	-4.7
IC4665	266.575	+05.717	350	~35	-0.5	-7
Collinder 359	270.275	+02.900	450	~100	-0.4	-8.
$\lambda$ -Ori	83.775	+09.933	450	3~5	0.8	-3.
$\sigma$ -Ori	84.675	-02.600	450	3~5	1.7	0.5
Cha I	166.700	-77.300	160~180	1~3	-21	1
Cha II	193.4	-77.2	160~180	1~3	-24	-8
Cha III	189.4142	-80.2533	160~180	1~3	?	?
NGC2516	119.51667	-60.75333	410	135	-4	11

Table 1.2: Secondary targets

Name	R.A. (J2000)	Dec (J2000)	Distance (pc)	Age (Myr)	$\mu_\alpha \cos(\delta)$ (mas/yr)	$\mu_\delta$ (mas/yr)
NGC 752	29.421	+37.785	460	1000	8	-12
NGC 6774	289.175	-16.283	~300	3000	-1	-29
$\alpha$ -Per	51.079	+49.862	171~200	50~70	22	-25
Praesepe	130.100	+19.667	600~700	1~3	?	?
NGC 2264	100.242	+09.895	920	3~5	-0.6	4.
M35	92.225	+24.333	850	~130	2.7	-3.5
M67	132.825	+11.800	860	3200~5000	-6.5	-4.5
NGC 6791	290.221	+37.772	4100	~4300	4.0	0.6
Coma Berenices	185.6262	+25.8450	96	~400	-12.4	-9.4
M39	322.950	+48.433	325	250	-8	-20
Trumpler 10	131.975	-42.450	425	180	-12	7
Collinder 285	220.22	+69.57	25	200	?	?
M41	101.504	-20.757	240	700	?	?
Stock2	33.679	+59.485	300	100	15	15
$\mu$ -Oph group	264.46129	-08.1188	120	-12	-21	
Upgren 1	188.75417	36.36667	110	?	-98	-49

The present work is the second generation of the DANCe intelligent systems (ISs). This new IS minimises the biases of the previous one and those present in most of the current classifiers in the literature. It steps ahead of the classification paradigm, the current one in the literature, and focuses on a new one: *learning the kinematic and photometric distributions of the cluster populations*, both single and binary stars. The classification task is now a by-product of this new IS. In the following section, I describe the properties of the previous DANCe IS classifier (the one from Sarro et al. 2014), together with the current ones from the literature.

## 1.4 Methodologies for the classification of cluster members

As mentioned above, the vast majority of the past and current methodologies for the analysis of stellar clusters work under the classification paradigm. Given the objects observables and the current knowledge of the cluster, they assign cluster membership probabilities. These are then used to classify the objects into the field or cluster populations. The classification depends on a probability threshold, which beside imposing a trade-off between completeness and contamination in the resulting sample, is not always established in an objective way.

In his pioneer work, which was also conducted in the Pleiades cluster, Trumpler (1921) selected candidate members using three criteria:

1. projected distance to the cluster centre,
2. similarity of the proper motions to those of the cluster, and
3. the degree of accordance with the relation between colour and brightness of the cluster.

During almost one century since the publication of Trumpler's work, his three criteria remain valid and are still widely used.

[Vasilevskis et al. \(1958\)](#) and [Sanders \(1971\)](#) put a solid probabilistic background to Trumpler's second criterium. [Vasilevskis et al. \(1958\)](#) expressed (in their Eq. 4) the probability of an object to belong to the cluster as

$$p_c = \frac{N_c \cdot \Phi_c(\lambda, \beta)}{N_c \cdot \Phi_c(\lambda, \beta) + N_f \cdot \Phi_f(\lambda, \beta)},$$

with  $N_c$  and  $N_f$  the number of cluster and field stars, respectively.  $\Phi_c$  and  $\Phi_f$  represent the bivariate normal distributions of cluster and field, respectively, and  $\lambda, \beta$  are the proper motions in galactic coordinates. The selection of galactic coordinates and the symmetry of the proper motion uncertainties allowed them to assume that the cluster bivariate normal distribution,  $\Phi_c$ , had circular symmetry (i.e. a diagonal covariance matrix). They fitted the means of these bivariate normal distribution by eye using histograms for each coordinate. [Sanders \(1971\)](#) used the same procedure of [Vasilevskis et al. \(1958\)](#), but fitted the parameters using the maximum-likelihood principle. This methodology remained almost intact during the following decades.

Later, [de Graeve \(1979\)](#) included the spatial information (Trumpler's first criterium) into the membership probability determination. Although his work is not available on-line, his formulation is given in the work of [Kozhurina-Platais et al. \(1995\)](#). In the latter, the cluster membership probability of an object, given its positions and proper motions, is

$$P_{\mu, \alpha, \delta} = \frac{N_c \cdot S_c(\alpha, \delta) \cdot \Phi_c(\mu_\alpha, \mu_\delta)}{N_c \cdot S_c(\alpha, \delta) \cdot \Phi_c(\mu_\alpha, \mu_\delta) + N_f \cdot S_f(\alpha, \delta) \cdot \Phi_f(\mu_\alpha, \mu_\delta)},$$

with  $\alpha$  and  $\delta$  the stellar positions in equatorial coordinates, and  $S_c$  and  $S_f$  the cluster and field spatial (positional) distributions. [Kozhurina-Platais et al. \(1995\)](#) assumed exponentially decaying and uniform distributions for the cluster and field spatial distributions, respectively. To model the proper motion distributions they also used bivariate normal distributions.

In the first decade of this century, authors started to incorporate Trumpler's third criterium in the selection of members. For example [Moraux et al. \(2003\)](#); [Deacon & Hambly \(2004\)](#); [Balaguer-Núñez et al. \(2007\)](#); [Lodieu et al. \(2012\)](#) used hard cuts in the colour, or in the colour magnitude diagrams (CMDs) to discriminate candidates. [Moraux et al. \(2003\)](#) used conservative cuts based on the theoretical isochrone models of [Baraffe et al. \(1998\)](#), while [Deacon & Hambly \(2004\)](#); [Balaguer-Núñez et al. \(2007\)](#); [Lodieu et al.](#)

([2012](#)) used arbitrary cuts in the CMDs diagrams. However, despite the conservative or objective these cuts may be, as pointed out by [Sarro et al. \(2014\)](#) they have at least two drawbacks. They render membership probabilities i) which are inconsistent across magnitude bins, and ii) only for those sources with measurements in the CMDs used for selection.

The previous works incorporated most of the observables (positions, proper motions and photometry) available for the vast majority of the stars. Other useful observables for the discrimination of cluster members are the parallax, the radial velocities and other indicators of youth (e.g. lithium abundance, photospheric activity, rotation, see for example [Barrado et al. 2016](#)). Although these observables can be of great use, they are only available for limited samples of objects. Furthermore, they tend to be biased towards the brighter objects. For these reasons, the members selection has been done using the astrometric and photometric observables: positions, proper motions and photometric bands.

In recent years, the astrophysical community started to develop automated methodologies for the cluster members selection. This development has been motivated by two main reasons. First, computing power has become widely available thus allowing the implementation of more complicated and computing demanding techniques. Second, the arrival of automated surveys providing incredible amounts of data for millions of objects have rendered obsolete the previous object-by-object oriented techniques. In the following, I will review the methodologies of [Malo et al. \(2013\)](#); [Gagné et al. \(2014\)](#); [Riedel et al. \(2017\)](#) for stellar associations, and the ones of [Krone-Martins & Moitinho \(2014\)](#); [Sarro et al. \(2014\)](#); [Sampedro & Alfaro \(2016\)](#) for star clusters. Although this work focuses only on the analysis of star clusters, the methodologies applied to associations are related and some times also applied to them. Thus, I review them as well.

[Malo et al. \(2013\)](#) develop a code for the Bayesian Analysis of Nearby AssociatioNs (BANYAN). It establishes membership probabilities for seven nearby young moving groups ( $\beta$  Pictoris, Tucana-Horologium, AB Doradus, Columba, Carina, TW Hydrea, and Argus) using a naive (no correlations included) Bayesian classifier. Their classification uses kinematic and photometric models of the seven moving groups and the field. These models are constructed with the previously known *bona fide* members and field objects. The photometric model in the low-mass range is extended beyond the known members using evolutionary models. One of the great advantages of this work is its ability to simultaneously obtain membership probabilities for the seven nearby young moving groups. It works on a six dimension observable space:  $I_c$  and  $J$  bands, proper motions and stellar positions. The recovery (True positive) and false alarm (False positive) rates of this classifier are 90% and 10%, respectively; the latter varying according to the moving group.

[Gagné et al. \(2014\)](#) improve [Malo et al. \(2013\)](#) methodology by developing BANYAN-II. It is tuned to identify low-mass stars and brown dwarfs by the use of redder photometric colours from *2MASS* and *WISE*. The spatial and kinematic distributions are modelled

with no alignment to the galactic coordinates as is done by BANYAN. They also included the observational uncertainties and the ability to deal with missing values, at least in parallax and radial velocities. When the parallaxes and radial velocities are not present, the estimated contamination rate ranges between 20% and 80% for a wide range (0.2-0.8) of probability cuts (see their Figure 5).

[Riedel et al. \(2017\)](#) developed a kinematic membership analysis code for LocAting Constituent mEmbers In Nearby Groups (LACeWING). Using only the spatial and kinematic information of stars, LACeWING, computes membership probabilities to 13 nearby young moving groups (NYMGs). LACeWING computes what they call the goodness-of-fit value, which is the distance between each object observables and the value of the NYMG divided by the quadratic sum of the uncertainties. Then it draws synthetic stars, compute their goodness-of-fit values, and bin them. The membership probability of a real object correspond to the fraction of synthetic stars, in the same goodness-of-fit bin, that were generated as members of the moving group. These authors do not provide estimates of contamination and recovery rates on synthetic samples. Instead, they directly give the number of recovered and false positives returned by LACeWING in each NYMG. As mentioned by these authors, the recovery rates are similar to those of BANYAN and BANYAN-II.

[Krone-Martins & Moitinho \(2014\)](#) created the Unsupervised Photometric Membership Assignment in Stellar cluster code (UPMASK). It establishes cluster membership probabilities in a frequentist approach using positions and photometry. It is an unsupervised and data driven iterative algorithm. Their methodology relies on clustering algorithms and the principal component analysis (PCA). It establishes membership probabilities by randomly generating synthetic stars based on the individual uncertainties. The membership of an object is the fraction of randomly generated stars in which the object was classified as a cluster member. Although untested, the authors mention that their methodology is able to incorporate proper motions, deal with missing data and with different uncertainty distributions. Measured on synthetic samples (with typical sizes ranging from thousands to tens of thousand objects), these authors find that UPMASK delivers true positive (recovery) rates grater than 80% for a probability classification threshold of  $p=0.9$  (they do not provide any reason for the use of this particular value). One of the great advantages of UPMASK is its ability to deal with superimposed clusters.

[Sarro et al. \(2014\)](#) created a Bayesian classifier to infer posterior membership probabilities based only on proper motions and photometry. Their cluster model is data driven and includes some parametric correlations (it is not a naive classifier like BANYAN or BANYAN-II). They construct the proper motions and photometric models of the field and cluster with clustering algorithms. However, the photometric model of the cluster is constructed using the principal curve analysis. Their treatment of uncertainties and missing values is consistent across observed features. They infer the best parameter values

of both cluster and field models using a Maximum-Likelihood-Estimator (MLE) algorithm, but based only on completely observed objects (i.e. without missing entries). Afterwards, they assign membership probabilities to objects with missing entries. They found an optimal probability classification threshold at  $p = 0.85$  using synthetic data sets of 2 million sources. At this threshold they measure a contamination rate of 4.5% and a true positive rate of 92.9%.

[Sampedro & Alfaro \(2016\)](#) created a Bayesian algorithm to estimate membership probabilities to open clusters. They compute the euclidean distance, in a multidimensional normalised space, of each object to the over-density of the cluster. Then, they assign membership probabilities according to this distance assuming univariate normal distribution for both cluster and field. Their methodology assumes that the cluster members are more densely concentrated than the field population. Since they use an euclidean metric to compute distances in a normalised space (normalised by the individual uncertainties), their methodology is highly affected by the heteroscedasticity of the presence of missing values in the observables. In an euclidean metric, objects with missing values have smaller, or at most equal, distances than those of objects with complete observations. They use a 0.5 probability classification threshold based on the Bayes minimum error rate, however, they do not provide error rates. On synthetic data sets of 500 objects they measure completeness and misclassification as a function of the similarity between the cluster and field univariate normal distributions. They do this varying i) the fraction of cluster members in the total data set size (from 20% to 80%), and ii) the number of variables (from among positions and proper motions). Their misclassification is in the range of 25% to 5% for two to four variables, while their completeness is better than 90%.

The previous methodologies perform well on the classification task, and successfully led to the identification of many new high probability members of nearby clusters and associations. However, key aspects still need to be tackled. These are

- Uncertainty of membership probabilities. These, as any other measurement, are uncertain. The previous authors report just the first moment of the object membership probability distributions.
- Treatment of objects with missing entries in their observables. In astronomy, measurements are strongly affected by the brightness of the source. The physical limits imposed by the dynamical range of detectors lead to non-uniform distributions of objects with missing entries. At the bright end, the excess of photons saturates the detector pixels and, most of the time, renders the measurements useless (however, see [Maíz-Apellániz 2003](#); [Olivares et al. 2013](#), for examples of high-precision astrometry and photometry on saturated images). At the faint end, sources with brightness below the limit of sensitivity are not detected. Furthermore, in a multi-wavelength data set the intrinsic stellar colours add another level of correlation between the sensitivity limits of different bands. The previous effects create a non-uniform

distribution of objects with missing values. In addition, missing values can also appear due to other non-random effects (e.g. proximity to a bright source), which will further affect any analysis that discards them. If the distribution of objects with missing entries will be uniform, then results based only on the completely observed (non-missing value) objects would be unbiased. However, the less uniform the distribution of missing-value objects is, the most biased the results are. See Chap. 8 of [Gelman et al. \(2013\)](#) for a discussion on missing at random and ignorability of the missing pattern. Objects with missing values have an important effect in the resulting samples of members. The cluster and field model are conditioned on the objects used to construct them. Since objects with missing values are not included in these models, the resulting samples of members may be biased.

- Heteroscedastic<sup>4</sup> uncertainties. Each object in the cosmos is unique and distinct from others. Assuming that all of them are alike is useful in certain situations, though. However, the assumption of homoscedasticity in the uncertainties of a sample, which is often done for practical reasons, may lead to biases. For example, it is known that the performances of PCA (used by [Krone-Martins & Moitinho 2014](#)), and consequently of the Principal Curve Analysis (used by [Sarro et al. 2014](#)), are strongly affected by heteroscedastic uncertainties ([Hong et al. 2016](#)).
- Sampling bias. Using a probability classification threshold to select candidate members results in a trade-off between contamination and completeness. To keep contamination under reasonable levels samples are never complete. Therefore, any further statistical analysis resulting from them suffers from a sampling bias.
- Propagation of uncertainties. Uncertainties in the observables must be propagated until the very end of subsequent astrophysical analyses. Currently, once a list of candidate members have been obtained, luminosity functions and then mass functions are derived on the basis of poisson uncertainties. At best, the uncertainties include the photometric ones, but completely discard proper motion uncertainties and those of the model that generated the list of candidates.

The methodology presented in this work attempts to address the previous problems.

## 1.5 A new intelligent system: the Bayesian Hierarchical Model

As mentioned before, this new IS steps ahead of the classification paradigm and focuses on a new one: *learning the kinematic and photometric distributions of the cluster*

---

<sup>4</sup>In statistics, homoscedasticity is the property of random sample of variables whose variance is homogeneous. This is, any sub-population shows the same variance as that of the population. By contrast, heteroscedasticity is the absence of homoscedasticity.

*populations*, both single and binary stars. This new IS addresses the aforementioned issues in the following way.

- Uncertainty of membership probabilities. Together with the kinematic and photometric distributions of the cluster population, it renders the cluster and equal-mass binaries (full) membership probability distributions, not just point estimates of them. Thanks to its Bayesian framework, it computes the probability distribution of the class, cluster vs field and single vs binary, given the object datum.
- Treatment of objects with missing entries in their observables. Thanks to its Bayesian framework, the new IS is able to include in its data set objects with missing entries in its observables. The missing values are treated as parameters which are marginalised with the aid of a prior.
- Heteroscedastic uncertainties. In the new IS, it is assumed that objects uncertainties share the same family distribution, the multivariate normal. However, the values of the uncertainties are not assumed to be the same. The new IS incorporates the intrinsic heteroscedasticity of the data. It assumes that the observed data results from the addition of a noise process to the *true*<sup>5</sup> data. The true intrinsic underlying relation is accessible after deconvolution of the observed uncertainties (see [Bovy et al. 2009](#), for another example of deconvolution).
- Sampling bias. The new IS avoids the bias associated to the use of only the high membership probability objects in the following way. It derives the kinematic and photometric distributions of the cluster populations taking into account each data set object proportionally to its cluster likelihood. Thus, there is no need of a probability classification threshold to obtain the clusters distributions.
- Propagation of uncertainties. Thanks to its Bayesian framework and the use of heteroscedastic uncertainties, the new IS is able to propagate the observational uncertainties directly into the posterior distribution of the parameters of the cluster model. Then, the distributions of the parameters can be propagated into the luminosity and mass distributions. The resulting mass distribution incorporates not just the observational uncertainties of all the objects, but also the uncertainties associated with the cluster model.

The core of this new IS is the Bayesian framework. The latter demands the use of prior distributions for the value of its parameters. To avoid as much as possible the subjectivity of choosing them, the new IS uses the Bayesian Hierarchical Model (BHM) methodology together with weakly informative priors ([Gelman 2006; Gelman et al. 2008; Huang & Wand 2013; Chung et al. 2015](#), see). The works of [Jefferys et al. \(2007\)](#); [Shkedy et al.](#)

---

<sup>5</sup>The true data is that which would be observed under negligible uncertainties.

(2007); Hogg et al. (2010b); Sale (2012); Feeney et al. (2013) are some examples of the use of BHMs in astrophysics. In the BHMs, the parameters of the prior distributions are given by other distributions in a hierarchical fashion. These parameters are also fitted from the data. Thus, BHM provide the most objective approach to the settling of prior distributions (Gelman 2006). However, it comes at a price: BHMs are computationally more expensive because they require far more parameters than standard approaches.

In such a way, the new IS learns the posterior distributions of the parameters in the BHM given the kinematic and photometric data. Due to the high dimension of the parametric space (the BHM has 85 parameters), the learning process demands a fast and reliable technique to obtain the posterior distributions of the parameters. Furthermore, the typical data set size for a cluster and the surrounding field is in the order of millions of sources. Thus the likelihood of the data must loop over all these millions of sources. The high dimensionality of the parametric space and the large data set size demand a technique that, on top of being fast and reliable, will also be compatible with parallel computing. Thus reducing to a minimum the computing time. The Markov Chain Monte Carlo techniques offer the demanded requirements. In specific, the *emcee* algorithm (Foreman-Mackey et al. 2013) was chosen due to its excellent performance properties and its ability to work in a parallel computing scheme.

The BHM presented in this work has been benchmarked on the Pleiades cluster, with the DANCe Data Release 2 (Bouy et al. 2015, DR2,) particularly. This reason roots in the following points. First, the Pleiades cluster together with the Orion Nebula Cluster are probably the most studied clusters in the history of astronomy. Thus, the current knowledge of Pleiades makes it ideal to benchmark any new tool or theory. Second, currently the DANCe project has processed the data of only three clusters: M35, CygOB2 and the Pleiades. CygOB2 and M35 are relatively poorly understood compared to the Pleiades, and are far from the sun (see Tables 1.1 and 1.2), which give them small proper motions values. Since the Pleiades is a well understood cluster, with well characterised lists of members (e.g Stauffer et al. 2007; Lodieu et al. 2012; Sarro et al. 2014; Bouy et al. 2015) and precise and accurate measurements, it is the perfect case study for the development, testing and validating of the new IS described here.

The rest of this work is structured as follows. In Chapter 2, I do a compilation of the current knowledge on the Pleiades cluster. In particular, that concerning its distance, spatial, velocity, luminosity and mass distributions. Also I give details of the Pleiades DANCe DR2 data set. The incredibly detailed knowledge of the Pleiades cluster and the high precision and carefully designed observations of the DANCe project are the main reasons for benchmarking the BHM on this cluster. Later, in Chapter 3, I give a brief introduction to probability theory and also present the details of the BHM methodology and of the techniques used to sample the statistical distributions of the cluster populations. Later, in Chapter 4, I give details of the analysis performed on synthetic data, of the

comparison of the recovered candidate members with those from the literature, of the kinematic and photometric distributions, and of the present day system mass distribution. Finally, in Chapter 5, I present my conclusions and my point of view concerning the work that must be done in the near future to continue improving our knowledge of the star formation process.

# Chapter 2

## The Pleiades as a benchmark

### 2.1 Generalities

The ancient greeks named Pleiades to a crowded group of nine stars which they believed shared a common origin. These stars were the seven sisters, that together with their parents the titan Atlas and the nymph Pleione, were put in the sky by the god Zeus.

Today, we call the Pleiades cluster not just to the nine stars that made up the original Pleione family, but to a much larger group, which according to [Bouy et al. \(2015\)](#) adds up to  $\sim 2100$  members. This cluster is fairly close to the sun,  $\sim 134$  pc (with parallaxes of  $7.44 \pm 0.08$  and  $7.48 \pm 0.03$  according to [Galli et al. 2017](#); [Gaia Collaboration et al. 2017](#), respectively), and is also young in galactic scales, with only  $\sim 125$  Myr ([Stauffer et al. 1998](#)). Since it is located in the solar neighbourhood, it has a distinctive angular velocity in the plane of the sky: about  $-16 \text{ mas} \cdot \text{yr}^{-1}$  in right ascension and  $20 \text{ mas} \cdot \text{yr}^{-1}$  in declination. It has a metallicity near to the solar one ( $[\text{Fe}/\text{H}] \sim 0$ , [Takeda et al. 2017](#)). Also, it has an almost null extinction of  $A_v = 0.12$  mag ([Guthrie 1987](#)). These properties make the Pleiades one of the most studied clusters in the history of astronomy <sup>1</sup>. Thus making it also the perfect test case of the methodology developed in this work.

As stated in the previous Chapter, the objective of the present study is to obtain the statistical distributions of the distance, position, velocity, luminosity and mass of the Pleiades cluster. Thus, in the following sections I will describe the current knowledge of the Pleiades, concerning these astrophysical quantities.

### 2.2 The distance to the Pleiades

#### 2.2.1 Measuring distances

In astronomy, measuring distances is a complicated task. Techniques vary according to the distance scale that they aim to measure. The distance ladder is constructed from

---

<sup>1</sup>Probably just after the Orion complex.

smaller to larger distances. The first step in this ladder is the distance to the sun. After that, the distance to the planets and then to the stars. This works deals only with nearby clusters, thus I only focus on measuring distances to these objects.

The most direct way to measure distance to nearby stars is by means of the trigonometric parallax. This is the maximum relative angular displacement, with respect to the far distant stars, that an object suffers in the course of a year. **It is usually reported in milliarc seconds (mas).** The relative displacement results from the movement of the Earth (thus of the observer) on its orbit around the sun. The relative displacement is maximal when measurements are taken at opposed points in the earth's orbit, when they are separated by six months. If the parallax were measured in seconds of arc and with infinite precision, then the distance to the object would be obtained by simply inverting its parallax. By doing so, the distance will be measured in parsecs. This unit gets its name from parallax-second. Thus an object located at a distance of one parsec from the sun shows a parallax of one arc second. The further the object is, the smaller the parallax gets.

As any measurement, parallaxes have uncertainties, which usually represent, or are a proxy for, the width of the distribution. **The parallax distribution is also continuous and non-limited.**

When transforming parallaxes into distances we may be tempted to take a summary of the distribution, the mean for example, and just invert it to obtain the distance. This only holds if the summary corresponds to the true value (i.e. the statistic is unbiased). **The true value is that which would be observed in the presence of negligible uncertainties.** However, because measurements have uncertainties, which almost always are not negligible, the inversion of the parallax not always renders an unbiased estimate of the distance. **Assuming that the distribution of parallax measurements of an object is Gaussian, Lutz & Kelker (1973) found that the distance to the object can be reasonably recovered by just inverting the parallax if its relative uncertainty is below 0.15-0.20.** However, **the shape of the distribution of parallax measurements of an object (the second and higher order moments) plays also an important role.** Transforming the parallax distribution into that of the distance require more than a simple inversion.

Several authors have proposed different approaches to the problem of distance determination using parallaxes, see for example Lutz & Kelker (1973); Bailer-Jones (2015); Astraatmadja & Bailer-Jones (2016a,b). The proper way, as Bailer-Jones (2015) points out, consists of inferring the true distances given the observed parallaxes. For that, a prior on the distance must be established. The authors mentioned before describe three different kinds of priors and the methodology needed to infer the true distances. However, going into deeper detail is beyond the scope of this work.

Now, I focus on the particular case of the distance to the Pleiades. One of the first mea-

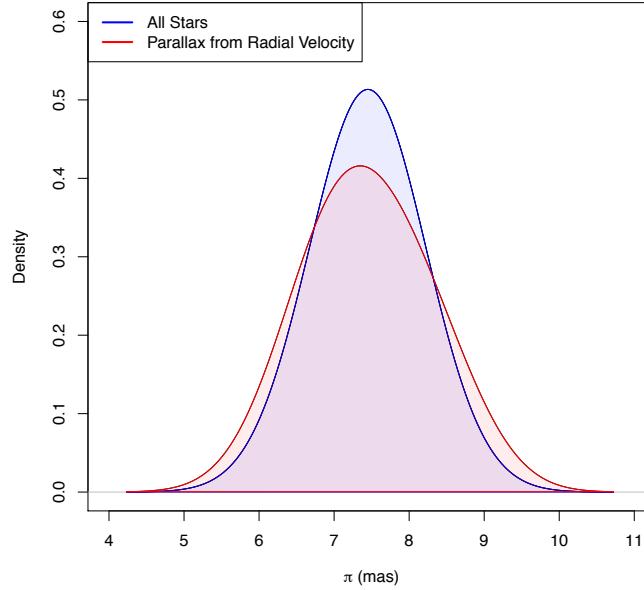
surements of the Pleiades distance using parallaxes was done by [van Leeuwen \(1999\)](#) using *Hipparcos* data. Later, the same author ([van Leeuwen 2009](#)) refined his analysis and obtained a value of  $120 \pm 1.9$  pc. However, [Gatewood et al. \(2000\)](#) with parallax measurements of seven stars taken at the Allegheny Observatory, and later [Soderblom et al. \(2005\)](#) with the parallaxes of three stars measured with the Fine Guidance Sensors of the *Hubble Space Telescope*, derived distances of  $130.9 \pm 7.4$  pc and  $134.6 \pm 3.1$  pc, respectively. Finally, [Melis et al. \(2014\)](#) using very long baseline radio interferometric parallaxes of three stars obtained a distance of  $136.2 \pm 1.2$  pc. There was a clear controversy between *Hipparcos* data and other parallax measurements. The current *Tycho-Gaia* data release (TGAS) of the *Gaia* mission, gives a distance to the Pleiades of  $133.7 \pm 0.5$  pc (from a parallax of  $7.48 \pm 0.03$  mas [Gaia Collaboration et al. 2017](#)). This seems to indicate that the *Hipparcos* parallaxes were somehow biased.

Our research group finds a distance to the Pleiades of  $134.4^{+2.9}_{-2.8}$  pc (from a parallax of  $7.44 \pm 0.08$  mas) ([Galli et al. 2017](#)), which is in good agreement with the one of TGAS. We found this distance using the kinematic parallaxes delivered by the moving cluster technique. This essentially exploits the fact that since clusters are bound, their members show a clear kinematic footprint: they seem to converge to a point in the sky ([Blaauw 1964](#)). Using this point and the velocity of the members (proper motion and radial velocities) it is possible to derive individual parallaxes. Furthermore, these individual parallaxes show a distribution which results from the dispersion of the cluster members distances along the line of sight. Figures [2.1a](#) and [2.1b](#) show the distribution of parallaxes for the Pleiades candidate members according to [Galli et al. \(2017\)](#) and [Gaia Collaboration et al. \(2017\)](#), respectively. As can be seen from these Figures, the results of both works agree on the mean of the parallax distribution. However, they recover different variances. This difference results from the discrepancy in the number of objects, 1210 in [Galli et al. \(2017\)](#) vs. 152 in [Gaia Collaboration et al. \(2017\)](#), and in the selection function of the two surveys. The TGAS sample is limited to the bright objects ( $V \sim 11.5$  mag), whereas the Pleiades DANCe DR2 includes the faint end of the distribution ( $i \sim 25$  mag). For these reasons, in the following I adopt the distance found by [Galli et al. \(2017\)](#).

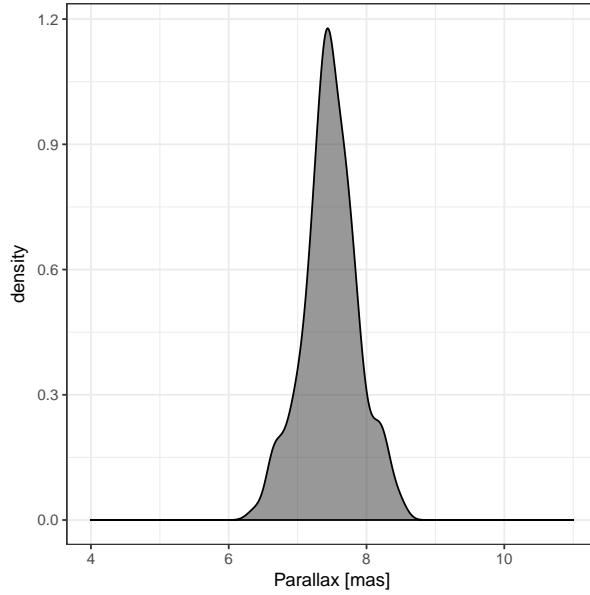
Nevertheless, the distance distribution (**measured uncertainties comprised**) is only the depth component of the space distribution of the cluster, the other two components are given by the projected spatial distribution.

## 2.3 Projected spatial distribution

The projected spatial distribution (PSD) is the two dimensional projection, in the plane of the sky (the one perpendicular to the line of sight), of the cluster three dimensional space distribution. In astronomy, object positions are



(a) Parallaxes according to [Galli et al. \(2017\)](#). The red line shows all their candidate members (1210) while the blue one only those with known radial velocity (64). Reproduced from Figure 13 of [Galli et al. \(2017\)](#), *A revised moving cluster distance to the Pleiades open cluster*, A&A, Vol. 598.



(b) Parallaxes according to [Gaia Collaboration et al. \(2017\)](#). Only their 152 candidate members.

Figure 2.1: Distribution of parallaxes for the Pleiades members.

commonly measured in what is called the Equatorial Coordinate System<sup>2</sup>. It can be thought as the projection of the geographical coordinates, latitude and longitude, into the sky. The Right Ascension (R.A.) coordinate, analogous to the longitude, gives the objects angle with the vernal equinox, in an eastward direction and along the celestial equator. The Declination (Dec.) coordinate, analogous to the latitude, gives the object angle perpendicular to the celestial

<sup>2</sup>Another common coordinate system is the Alt-Azimuth one (see [Smart & Green 1977](#)).

equator, positive to the North and negative to the South. For more details of the Equatorial Coordinate System see for example [Smart & Green \(1977\)](#).

Stellar positions of an object are far more easily measured than its parallax. For this reason, just a small fraction of objects with stellar positions has also parallax measurements. In the case of the Pleiades, after cross-matching the *Hipparcos* catalogue ([Perryman et al. 1997](#)) with the candidate members of [Bouy et al. \(2015\)](#), I find that only 70 of the  $\sim 2100$  candidates have parallaxes. As seen in the previous section, this figure is roughly doubled with the new *Gaia* DR1 ([Gaia Collaboration et al. 2016b](#)). In addition to the scarcity of parallaxes, which is expected to be solved with future *Gaia* data releases, the relative uncertainties in R.A. and Dec. coordinates, measured in degrees, are far better ( $\sim 10^{-5}$ ) than those of the parallaxes ( $\sim 10^{-1}$ ), measured in mas. Transforming these relative precisions into parsecs, by means of the distance, it is seen that the position in the plane of the sky is  $10^4$  more precise than along the line of sight. These two are the main reasons, for which the Pleiades space distribution has been studied mainly through its PSD. The latter has been the subject of several studies.

One of the earliest results of the Pleiades PSD was done by [Limber \(1962\)](#). He used a mixture of four indices polytropic distribution, as was described in his earlier [Limber \(1961\)](#) work, to fit the PSD of the 246 candidate members of [Trumpler \(1921\)](#). These candidates were contained in a  $3^\circ$  radius around *Alcyone* (one of the central most massive stars of the Pleiades cluster).

Later, [Pinfield et al. \(1998\)](#) fitted King profiles ([King 1962](#)) to candidate members from the literature, which were contained in a  $3^\circ$  radius area. They fitted King profiles to objects within different mass ranges, their bins centred at  $5.2, 1.65, 0.83$  and  $0.3 M_\odot$ . The tidal radius they found,  $13.1 \text{ pc}$  ( $\sim 5.6^\circ$ ) contained 1194 candidate members. The total mass of these members amounted to  $735 M_\odot$ . These authors also estimated a mean individual stellar mass of  $0.616 M_\odot$ . **They measured core radius in the 0.9 to 2.91 pc range for the King profiles fitted to their different mass bins.**

On the same year [Raboud & Mermilliod \(1998\)](#) also fitted a King's profile ([King 1962](#)) to a list of 270 candidate members with masses in the range  $0.74 - 7.04 M_\odot$ , which were contained within a  $5^\circ$  radius area. They found a core radius of 1.5 pc and a tidal radius of 17.5 pc (7.5 degrees). Using different approaches, they derived a total mass within the range of  $500 - 8000 M_\odot$ . They also measured an ellipticity of  $\epsilon = 0.17$ , however they did not make any explicit mention on the position angle of the axis of the ellipse.

Later, [Adams et al. \(2001\)](#) also fitted a King profile to objects with membership probabilities  $p > 0.3$  within a radius of  $10^\circ$ . They found a core radius of  $2.35 - 3.0$  pc and a tidal radius of  $13.6 - 16$  pc ( $5.8 - 6.8^\circ$ ). They estimate a total mass of  $\sim 800 M_\odot$ , and their measured ellipticities are in the range  $0.1 - 0.35$ .

[Converse & Stahler \(2008\)](#) fit a King profile to a sample of 1245 candidate members from [Stauffer et al. \(2007\)](#) compilation. These objects have masses greater than  $0.08 M_{\odot}$  and are contained within a  $5^{\circ}$  radius. They obtained a tidal radius of 18 pc (7.7 degrees) and a core radius of 1.3 pc. Later, [Converse & Stahler \(2010\)](#) refined their study and obtained a core radius of  $2.0 \pm 0.1$  pc, a tidal radius of  $19.5 \pm 1.0$  pc ( $\sim 8.3$  degrees) and a total mass of  $870 \pm 35 M_{\odot}$ . In Fig. 2.2a, I reproduce the surface density fit obtained by these authors.

The previous summary of results shows at least two interesting points. In the first place, King profile ([King 1962](#)) has been the preferred choice for the Pleiades cluster, although it was created to fit the PSD of globular clusters. Since globular clusters are farther away than open clusters and in a low density environment, usually the end of their PSD is well within the survey area. The second point concerns the increasing trend of the tidal radius with the size of the survey and the publication date, see Table 2.1. As the surveys increase in area the derived tidal radii increase as well. The exception is the work of [Adams et al. \(2001\)](#). Since these authors used low membership probability ( $\geq 0.3$ ) objects, they may have also fitted the field. The surface density of a tidally truncated cluster should diminish with radius and eventually go to zero at the tidal radius. However, as can be seen in Figure 2.2b, where I reproduce Figure 8 of [Adams et al. \(2001\)](#), their surface density remains almost constant after  $5^{\circ}$ . This may be an indication of contamination in their sample. Furthermore, as those authors mention, they expect that the contamination dominates their sample outside the  $5^{\circ}$  radius.

The two points mentioned before are tightly related. With the exception of the work of [Adams et al. \(2001\)](#), the coverages of the rest of the surveys have not reached their estimated tidal radius. It indicates that the sample of members we currently have is spatially biased. It only contains objects from the inner parts of the cluster. Thus, estimates of the tidal radius may also be biased. Nevertheless, this issue will be addressed with the full sky coverage of *Gaia*'s data.

Table 2.1: Survey, and derived core and tidal radius for recent studies in the literature.

Authors	Core radius (pc)	Tidal radius (pc)	Tidal radius ( $^{\circ}$ )	Survey radius ( $^{\circ}$ )
<a href="#">Pinfield et al. (1998)</a>	0.9-2.91	13.1	5.6	3
<a href="#">Raboud &amp; Mermilliod (1998)</a>	1.5	17.5	7.5	5
<a href="#">Adams et al. (2001)</a>	2.35-3.0	16	6.8	10
<a href="#">Converse &amp; Stahler (2008)</a>	1.3	18	7.7	5
<a href="#">Converse &amp; Stahler (2010)</a>	2.0	19.5	8.3	5

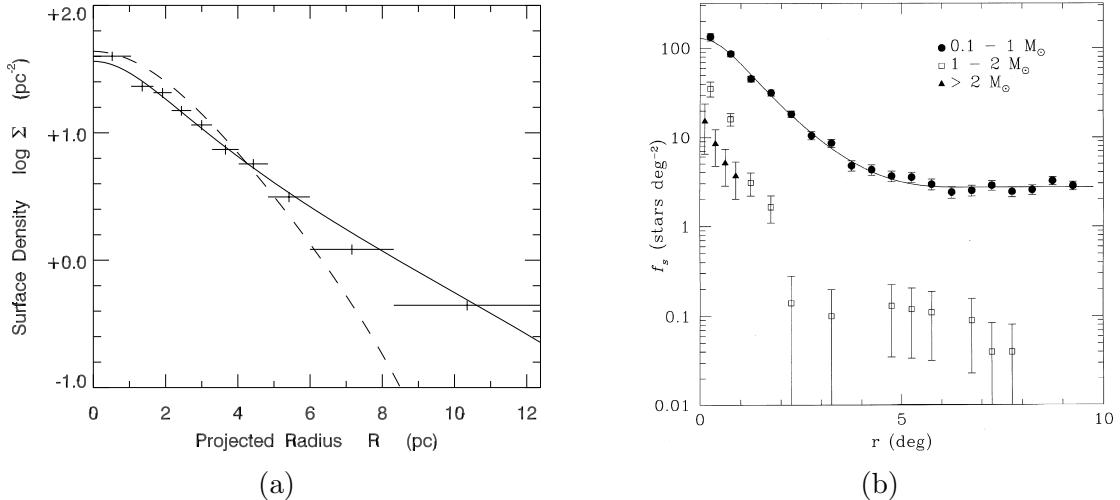


Figure 2.2: Projected spatial distribution of the Pleiades cluster. (a) Results from [Converse & Stahler \(2010\)](#). The crosses, and the dashed and solid lines represent the data and the fitted polytrope and King profile, respectively. Reproduced from Figure 1 of [Converse & Stahler \(2010\)](#), *The dynamical evolution of the Pleiades*, MNRAS, Vol. 405. (b) Results from [Adams et al. \(2001\)](#). The line shows the fitted King profile while the symbols are for different mass bins used. Reproduced from Figure 8 of [Adams et al. \(2001\)](#), *The Mass and Structure of the Pleiades Star Cluster from 2MASS*, AJ, Vol. 121.

## 2.4 Velocity Distribution

The three dimensional velocity distribution of the Pleiades has also been studied using its projections. One of them goes along the line of sight, it corresponds to the radial velocity. The other one is perpendicular to the previous one, lies in the plane of the sky, and corresponds to the transverse velocity. It is derived from proper motions. These are angular velocities obtained after measuring the angular displacement of the object in at least two different epochs. Again, measuring the individual stellar position and its displacement over time is far easier than measuring radial velocities. These are measured using the Doppler shifted absorption lines in the spectre of the object. This shift is proportional to the object velocity relative to the observer along the line of sight.

Since radial velocities require the object spectrum, their obtention for all cluster members, and particularly for the fainter ones, will demand a large amount of observing time. On the other hand, wide field images have been available for quite a long time, thus allowing long time base lines to measure proper motions. Nevertheless, due to the Pleiades distance, radial velocities are often more precise than proper motion measurements, usually on the  $1 \text{ km} \cdot \text{s}^{-1}$  regime. For these reasons, historically, the velocity distribution of the Pleiades cluster has been studied through the proper motions of its members.

Probably the first description of the transverse velocity distribution (TVD) of the Pleiades is that of Pritchard (1884). Using archival data from Königsberg (1838-1841), Paris(1874) and Oxford (1878-1880) observatories, together with his own *Differential Micrometer* observations, he was able to observe the relative displacements of 40 Pleiades

stars. According to him ([Pritchard 1884](#)): *the relative displacements of these distant suns, although not distinctly and accurately measurable in numerical extent, appear to vary both in direction and amount; indicating thereby the mutual influence of a group of gravitating bodies.*

Later, [Trumpler \(1921\)](#) used, for the first time, proper motion measurements to identify the members of the Pleiades cluster. He classified objects as candidate members according to the distance they show, in the proper motion space, to the mean proper motion of the cluster. This mean was previously calculated by Boss in his *Preliminary General Catalog* ([Boss 1910](#)). So far as my historic research went, Boss' work was the first measurement of a statistic of the TVD of the Pleiades.

Later [Titus \(1938\)](#), using Trumpler's data and archival compilations, was able to measure the dispersion of the proper motions distribution. He estimated it to be  $0.79 \text{ mas} \cdot \text{yr}^{-1}$  ( $0.65 \text{ km} \cdot \text{s}^{-1}$  at 136 pc). This was probably, the first measurement of the second moment of the spatial velocity distribution. From this value he then derived a total mass of  $260 \text{ M}_\odot$ .

In recent years [Pinfield et al. \(1998\)](#) used the velocity dispersion to probe that the cluster was in an state near to the virial equilibrium. Later, [Loktin \(2006\)](#) used the projected radial and tangential velocity components of the spatial velocity distribution of 340 members to claim the absence of evidence for rotation, expansion or compression of the cluster. Also, he also found no evidence to support mass segregation.

Concerning the radial velocities, the first record for the Pleiades correspond to [Adams \(1904\)](#). He measured the radial velocities of the six brightest stars. After this seminal work, more than a dozen of works have been published. Among them are the works of [Wells \(1924\)](#); [Smith & Struve \(1944\)](#); [Mermilliod \(1979\)](#); [Liu et al. \(1991\)](#); [Rosvick et al. \(1992\)](#); [Stauffer et al. \(1994\)](#); [Mermilliod et al. \(1997\)](#); [Martin et al. \(1996\)](#); [Terndrup et al. \(2000\)](#); [Loktin \(2006\)](#); [Wheeler \(2009\)](#); [Mermilliod et al. \(2009\)](#), and [Kordopatis et al. \(2013\)](#). In previous studies, the typical number of Pleiades candidate members was below 100 objects, with the works of [Wheeler \(2009\)](#) and [Mermilliod et al. \(2009\)](#) reaching 269 and 275 objects, respectively. The latest compilation of radial velocities from the literature is the one made made by [Galli et al. \(2017\)](#). This list contains measurements for 394 objects. The distribution of these radial velocities is almost gaussian with a centre at  $5.6 \text{ km} \cdot \text{s}^{-1}$ . In [Galli et al. \(2017\)](#), we estimated a velocity dispersion of the  $0.8 \text{ km} \cdot \text{s}^{-1}$  for the Pleiades candidate members.

Although, transverse and radial velocities are useful projections, the dynamical analysis of the cluster demands the three dimensional distribution. In [Galli et al. \(2017\)](#) we provide a list of 64 cluster members with full spatial velocities. The distributions of the three projections of these spatial velocities are shown in Figure 2.3.

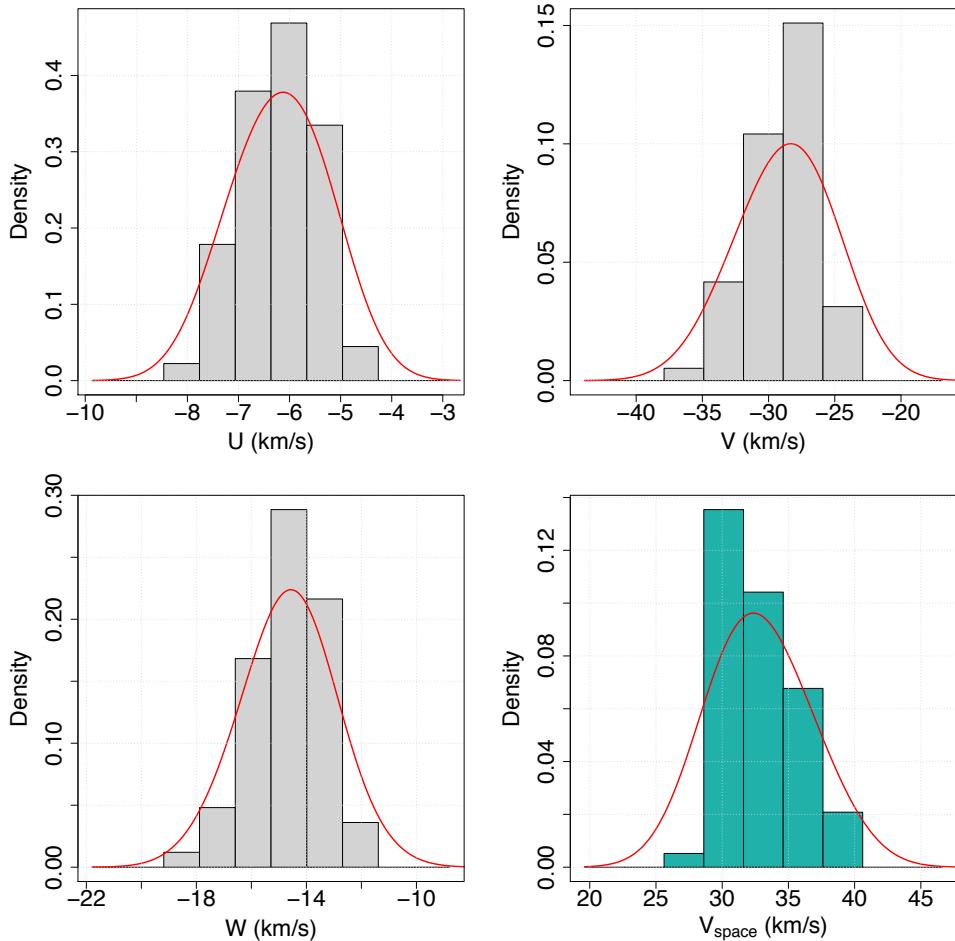


Figure 2.3: Histogram and kernel density estimation (red line) of the components (grey) and modulus (green) of the spatial velocity distribution of 64 candidate members of [Galli et al. \(2017\)](#) with radial velocities and parallaxes. Reproduced from Figure 11 of [Galli et al. \(2017\)](#), *A revised moving cluster distance to the Pleiades open cluster*, A&A, Vol. 598.

## 2.5 Luminosity Distribution

The luminosity distribution usually refers to the statistical probability distribution of the absolute<sup>3</sup> magnitude of the cluster population. It can also refer to the distribution of apparent magnitudes. It can be thought as the spectrum of brightness of the cluster members. Its importance lies in the fact the the luminosity of a star, measured in absolute magnitudes, can be related to its mass by the mass-luminosity relation. Therefore, the luminosity distribution is a proxy for the mass distribution.

The study of the distribution of luminosities in the Pleiades started few years later than those of the positions and proper motions. The first record I found on the luminosity distribution is the one of [Trumpler \(1921\)](#) (see Fig. 2.4). He computed the number of

---

<sup>3</sup>The absolute magnitude  $M$ , is the brightness that an object of apparent magnitude  $m$  will show at a distance of 10 pc.

stars in each magnitude bin for his two samples of candidate members, those comprising the objects within the central  $1^\circ$ , and those between  $1^\circ$  and  $3^\circ$ , referred as Tables I and II, respectively. The completeness of the inner and outer samples was estimated at 14.5 and 9.8 photographic magnitudes (roughly 14 and 9 in the visual band), respectively. He observed that the luminosity distributions of these two samples were not alike, with the inner sample being brighter than the outer one. He also noticed that the luminosity distribution is not smooth, and shows a local minimum at 9 magnitudes, then an abrupt rise. Both effects are present in the two samples.

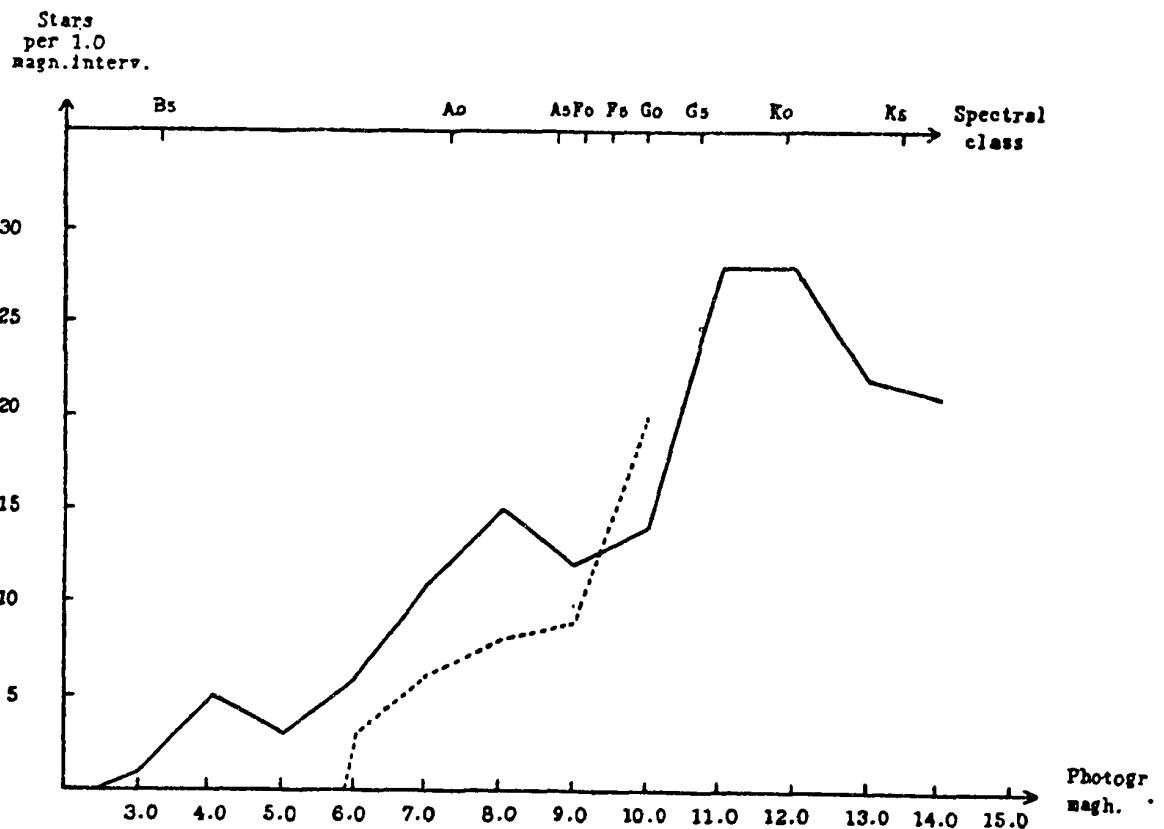


Figure 2.4: Luminosity distribution according to [Trumpler \(1921\)](#). The solid and dashed lines correspond to objects within  $1^\circ$  and within  $1^\circ$  and  $3^\circ$  from the centre. Reproduced from Figure 2 of [Trumpler \(1921\)](#), *The physical members of the Pleiades group*, Lick Observatory Bulletin, Vol. 10.

Later, [Johnson & Mitchell \(1958\)](#) obtained the luminosity distribution using a sample of 289 candidate members. They assessed membership solely on photometry. Their luminosity distribution is shown in Fig. 2.5

Later, [Limber \(1962\)](#) compared the luminosity distributions derived from the data of [Trumpler \(1921\)](#), [Hertzsprung \(1947\)](#), and [Johnson & Mitchell \(1958\)](#), with the initial luminosity distribution that he derived ([Limber 1960](#)). The initial luminosity distribution corresponds to the distribution of luminosities that the cluster had at the moment of formation. [Limber \(1960\)](#) derived it mixing data of galactic clusters and the local neigh-

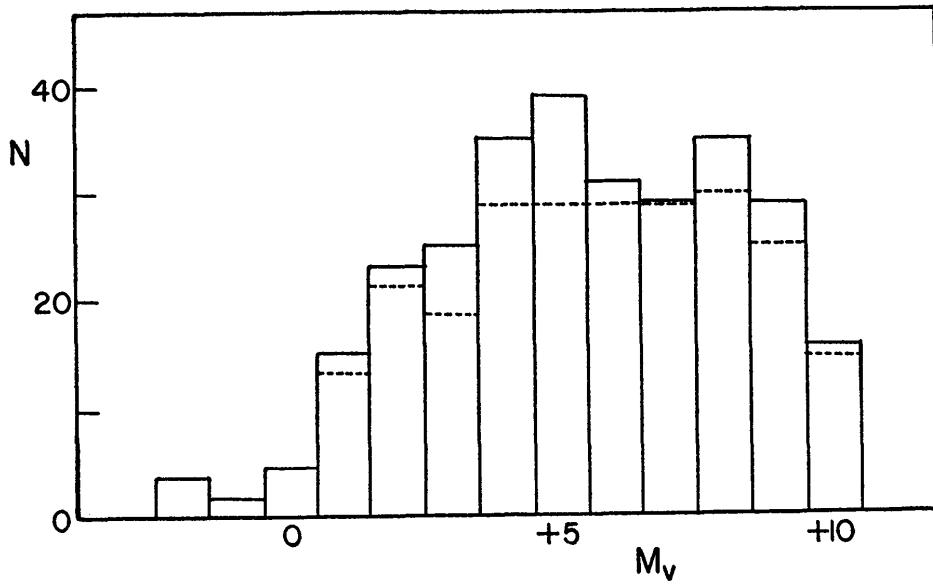


Figure 2.5: Luminosity distribution in the visual band according to [Johnson & Mitchell \(1958\)](#). The dotted line represent the counts of main sequence stars only. Reproduced from Figure 3 of [Johnson & Mitchell \(1958\)](#), *The Color-Magnitude Diagram of the Pleiades Cluster. II.*, ApJ, Vol. 128.

bourhood, and later correcting it by effects of age. He noted that the Pleiades present day luminosity distribution starts to differ from the initial luminosity distribution at visual magnitude 5.5, see Fig. 2.6. Assuming that this difference is due to the fact that stars fainter than 5.5 have not yet had enough time for contraction, he derives an age of 50 Myr.

In recent years, the luminosity distribution has been described in the works of [Lodieu et al. \(2012\)](#) and [Bouy et al. \(2015\)](#). [Lodieu et al. \(2012\)](#), using the *UKIDSS* DR9 survey for galactic clusters and a probabilistic membership selection method (see discussion in Chapter 1) based on proper motions only, and proper motions and photometry, found 8797 and 1147 candidate members, respectively. However, they do not provide the contamination rate in their analysis. Using both lists, they provide their luminosity distributions in the  $Z$  band, which I show in Fig. 2.7.

In [Bouy et al. \(2015\)](#), we estimated the present day system luminosity distribution of 1378 candidate members contained within the central  $3^\circ$  region (with the centre at  $RA = 03 : 46 : 48$  and  $Dec = 24 : 10 : 17$  J2000.0). It is called systemic because it has not been corrected for unresolved systems. An unresolved system is a group of stars (e.g. binaries) that due to its compactness appears as a single object. This distribution was computed for the  $K_s$  band and is sensitive up to  $K_s \sim 20$  mag and complete until  $K_s \sim 17$  mag. This luminosity distribution is reproduced in Fig. 2.8

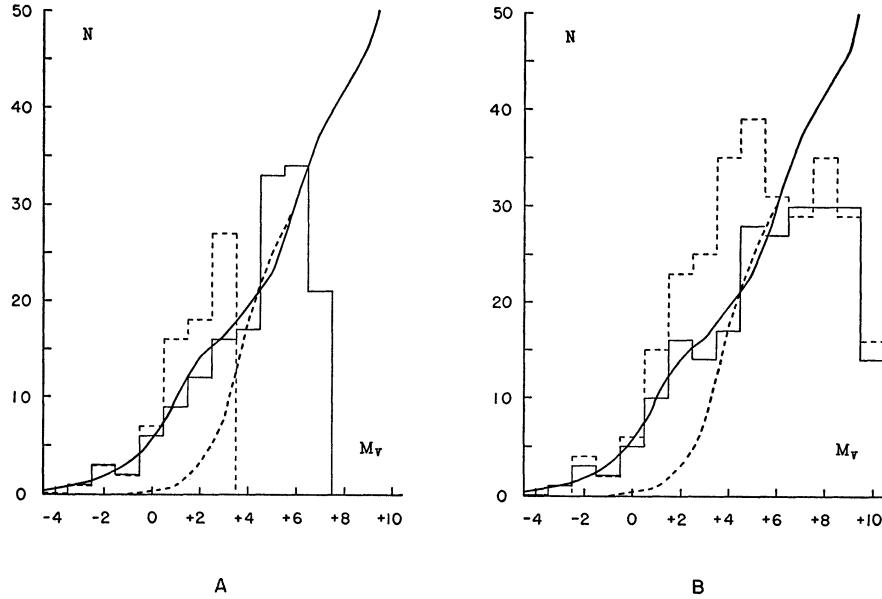


Figure 2.6: Luminosity distribution in the visual band according to Limber (1962). The solid and dashed histograms in: A correspond to Trumpler (1921) data from the Tables II and I, respectively, in B, correspond to the data from Hertzsprung (1947) and Johnson & Mitchell (1958), respectively. The solid and dashed curves line represent initial luminosity distribution and the present day luminosity distribution of the solar neighbourhood, respectively, both from Limber (1960) . Reproduced from Figure 4 of Limber (1962), *The Dynamics of the Pleiades Cluster.*, ApJ, Vol. 135.

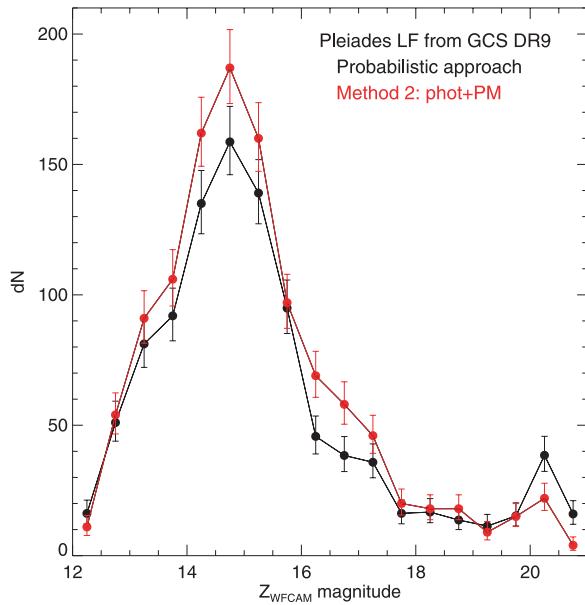


Figure 2.7: Luminosity distribution in the  $Z$  band according to Lodieu et al. (2012). The red and black lines correspond to the two different probabilistic methods. Reproduced from Figure 9 of Lodieu et al. (2012), *Astrometric and photometric initial mass functions from the UKIDSS Galactic Clusters Survey - I. The Pleiades*, MNRAS, Vol. 422.

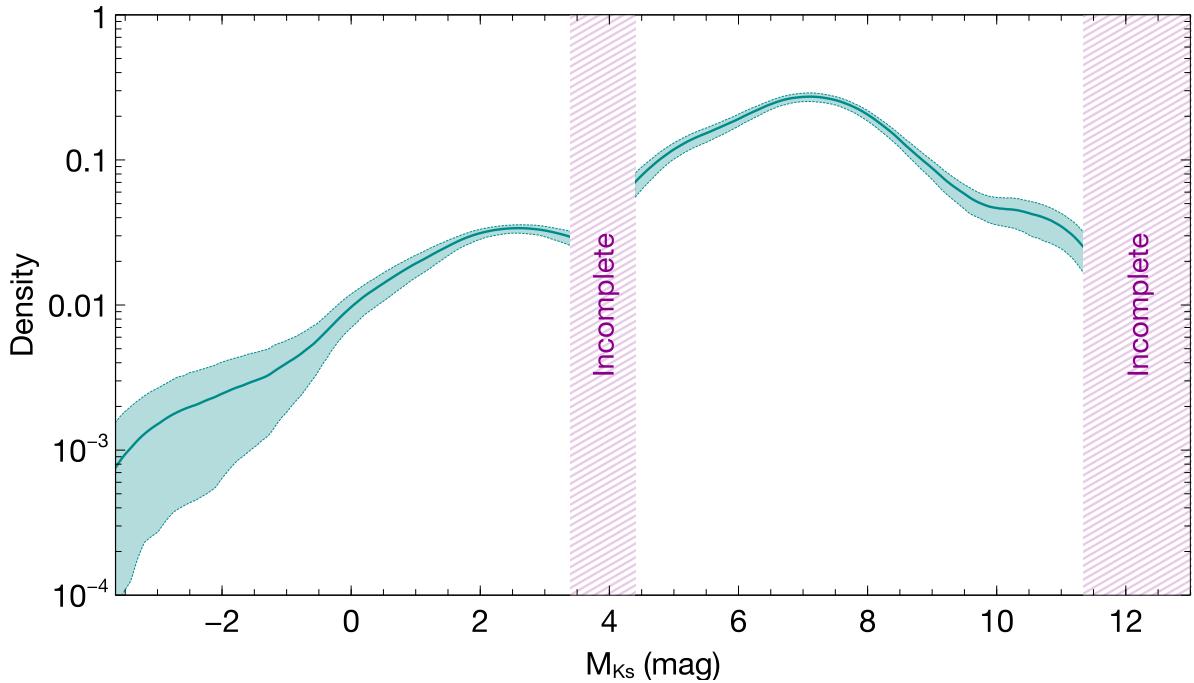


Figure 2.8: Luminosity distribution in the  $K_s$  band according to [Bouy et al. \(2015\)](#). The incompleteness regions are shaded. Reproduced from Figure 8 of [Bouy et al. \(2015\)](#), *The Seven Sisters DANCe. I. Empirical isochrones, luminosity, and mass functions of the Pleiades cluster*, A&A, Vol. 577.

## 2.6 Mass Distribution

In astrophysics, the mass distribution is a cornerstone in the understanding of the star formation process and the later evolution of stellar systems. Although the temporal evolution of these systems is mainly dominated by the gravitational potential, the initial conditions and an ongoing star formation process, if any, can contribute to the shape of the mass distribution. This last contains the fingerprints of past events in the history of the cluster and plays a key role in its future evolution. Indeed, the evolution of the mass distribution is an essential element in one of modern astrophysics' objectives: the determination of the role played by the initial conditions or the environment, in the temporal evolution of the stellar systems. The mass distribution at the moment of the cluster formation, which is known as the initial mass distribution evolves in time according to: i) stellar internal and atmospheric processes (e.g. contraction, mass loss, inflation, supernova events), ii) population dynamical interactions (e.g. three-body encounters, runaway stars, stellar evaporation), and iii) galactic dynamics (e.g. tidal effects, encounters with other clusters). For these reasons, the study of the initial mass function and its posterior evolution is a key element in the current understanding not just of open clusters but also of galactic and extragalactic populations.

The mass distribution of the Pleiades has been largely studied. The first work on the mass distribution is that of [Limber \(1962\)](#). Although he did not show any graphical or

tabular representation of it, he gave the luminosity distribution and the mass-luminosity ratio. From these the mass distributions can be derived. Instead, he used them to obtain the total mass of the cluster ( $760 M_{\odot}$ , see next Section).

Most probably, the first work to present the mass distribution derived from luminosity distributions and a mass-luminosity relation from theoretical models was that of [Hambly & Jameson \(1991\)](#). Using  $R$  and  $I$  observations from the *United Kingdom Schmidt Telescope Unit* together with the mass-luminosity relation from theoretical isochrone models of Padova group, he was able to transform his luminosity distribution into a mass distribution. In Fig. 2.9, I reproduce his results.

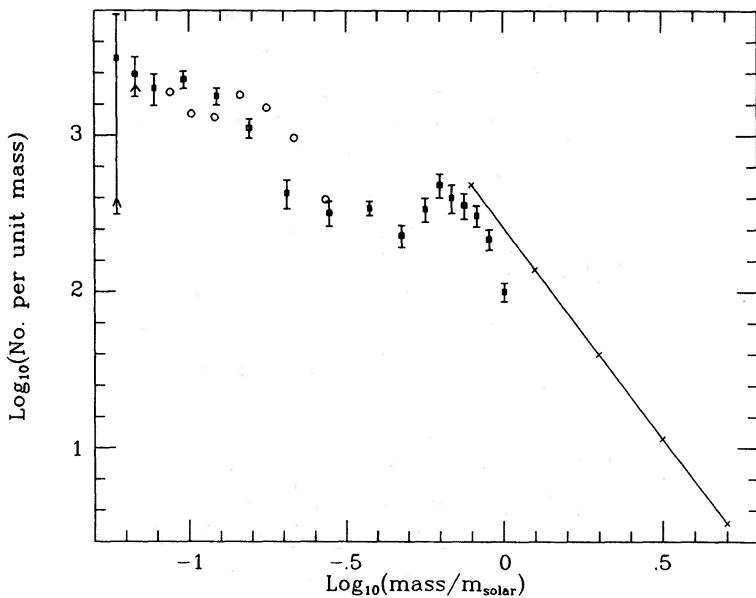


Figure 2.9: Mass distribution of [Hambly & Jameson \(1991\)](#) derived from luminosity distribution and theoretical isochrone models. The open circles result from assuming an older age of 200 Myr. The line represents the mass distribution of [van Leeuwen \(1980\)](#). Reproduced from Figure 11 of [Hambly & Jameson \(1991\)](#), *The luminosity and mass functions of the Pleiades - Low-mass stars and brown dwarfs*, MNRAS, Vol. 249.

From the year 2000 till date, several studies have been published in which the subject of analysis is the Pleiades mass distribution, e.g. [Hodgkin & Jameson \(2000\)](#); [Jameson et al. \(2002\)](#); [Moraux et al. \(2001, 2003, 2004\)](#); [Lodieu et al. \(2007\)](#). However, for the sake of simplicity, here I only analyse the two most recent works, those of [Lodieu et al. \(2012\)](#) and [Bouy et al. \(2015\)](#). The Present-Day System Mass Distributions (PDSMD) derived from both these works are shown in Figs. 2.10 and 2.11. These works obtained first the luminosity distribution, and then transformed it into a mass distributions using mass-luminosity relations of theoretical isochrone models.

[Lodieu et al. \(2012\)](#) used a distance of 120.2 pc, an age of 120 Myr, and the *NEXTGEN* theoretical models of [Baraffe et al. \(1998\)](#) to transform the luminosity into the mass distribution. On the other hand, in [Bouy et al. \(2015\)](#) we use a distance of 136.2 pc and

age of 120 Myr and the *BT-Settl* theoretical isochrone models of Allard (2014).

Both works found that their derived PDSMD are in general agreement with the IMF of Chabrier (2005) for unresolved systems (see Section 1.1). However, in Bouy et al. (2015) we pointed out that, assuming that the theoretical mass-luminosity relation is correct, the IMFs of Chabrier (2005) and Thies & Kroupa (2007) predict too many low-mass stars and brown dwarfs in the range  $0.04 - 0.1 M_{\odot}$ .

The differences at the low mass regime between the PDSMD derived by Lodieu et al. (2012) and Bouy et al. (2015) (see Figs. 2.10 and 2.11) could arise from the different i) samples of members, ii) mass-luminosity relations, and iii) distances adopted by both works.

The different distances can introduce a general shift in the luminosity, which in turn can shift the mass distribution. This shift in luminosity, however, can be neglected due to its small value, 0.06 mag.

Concerning the differences between the mass-luminosity relations of the two isochrone models, Allard et al. (2013) show that there are clear differences between the effective temperatures delivered by the *BT-Settl* and the *NEXTGEN* models in the low-mass regime, at 5 Gyr particularly. **However, they do not discuss how this difference change in younger ages. Thus, this effect can not be discarded as the source of the differences.**

Concerning the differences between the lists of candidate members, Lodieu et al. (2012) do not provide (at least explicitly) any estimate of contamination rate of their samples. Furthermore, their membership methodology has some draw backs (see Sarro et al. 2014) that may have biased their results. **Therefore, the agreement that Lodieu et al. (2012) found between their present day mass distribution and the IMF of Chabrier (2005), which models the field mass distribution, may be an indication that their sample of candidate members is contaminated by the field.**

On the other hand, in Bouy et al. (2015) we estimated a contamination rate of 7%. However, we have no evidence for it to be non-homogeneous in mass. Even if the 7% contaminants were not homogeneously distributed in the mass range, this value is not able to account for the observed discrepancies (30 – 40% in the low-mass regime) between the IMF of Chabrier (2005) and our present day mass distribution.

The previous studies show that there is still work to do in the analysis of the Pleiades mass distribution, particularly at the low-mass regime where the IMFs show discrepancies with the observed present day mass distribution.

### 2.6.1 Total mass of the cluster

Before ending this section I present a (non exhaustive) summary of the studies that provided an estimate of the total mass of the cluster.

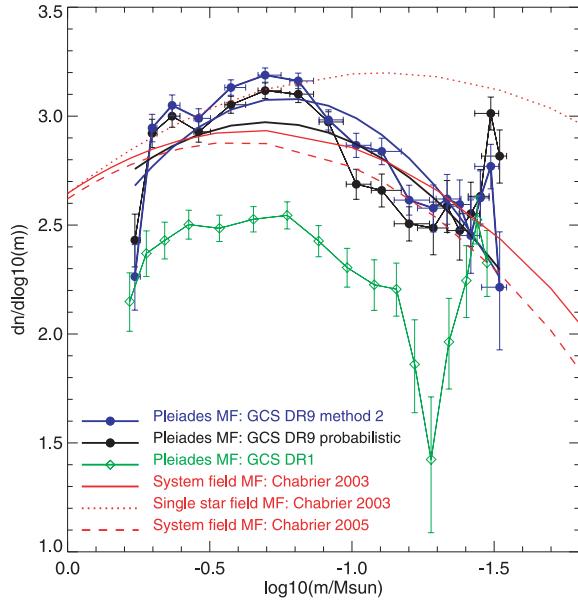


Figure 2.10: Pleiades present day mass distribution from [Lodieu et al. \(2012\)](#). GCS stands for Galactic Cluster Survey. The first and last two points must be treated with caution due to saturation and contamination at the bright and faint ends, respectively. Reproduced from Figure 9 of [Lodieu et al. \(2012\)](#), *Astrometric and photometric initial mass functions from the UKIDSS Galactic Clusters Survey - I. The Pleiades*, MNRAS, Vol. 422.

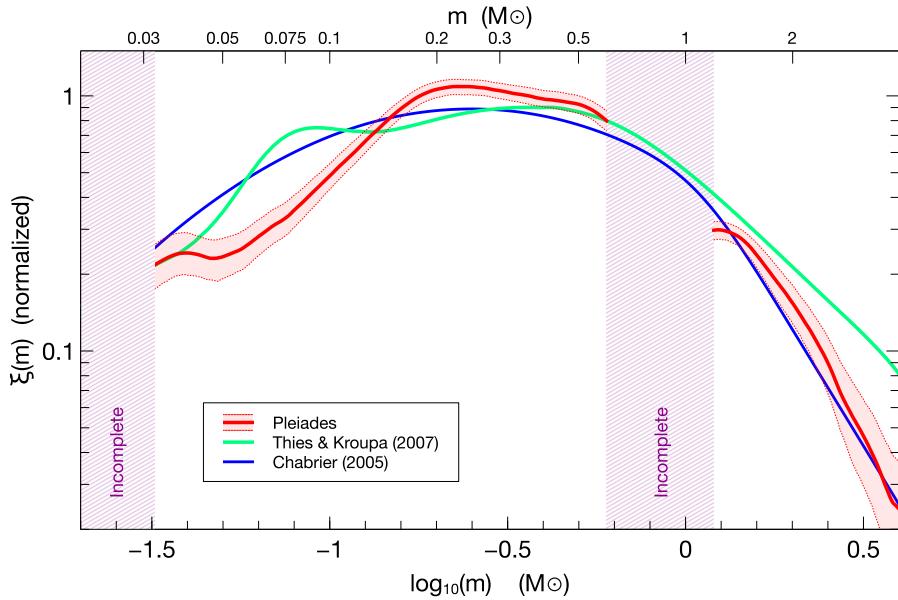


Figure 2.11: Pleiades present day mass distribution from [Bouy et al. \(2015\)](#) (red). IMFs from [Chabrier \(2005\)](#)(blue) and [Thies & Kroupa \(2007\)](#) (green) are also shown. Reproduced from Figure 9 of [Bouy et al. \(2015\)](#), *The Seven Sisters DANCE. I. Empirical isochrones, luminosity, and mass functions of the Pleiades cluster*, A&A, Vol. 577.

The first record I found of the cluster total mass is that of [Titus \(1938\)](#). He estimated a total mass of  $260 M_{\odot}$  assuming virial equilibrium. He also computed  $200 M_{\odot}$  using the

Eddington's mass-luminosity relation for objects brighter than 15 mag in the visual band.

The subsequent works continue to report higher masses. Woolley (1956) estimated a total mass of  $337 M_{\odot}$  using a polytrope model fitted to Hertzsprung's catalogue. He then mentions that taking into account Trumpler's data, the total mass should be about  $500 M_{\odot}$ .

Limber (1962) computed the total mass in two ways. In the first one he assumed the cluster was virialised and obtained a mass of  $900 M_{\odot}$ . Using the luminosity function he estimated the lower limit to the total mass in  $760 M_{\odot}$ .

Jones (1970) measured  $470 M_{\odot}$  and  $690 M_{\odot}$  using the luminosity distribution and the virial theorem, respectively.

Later, van Leeuwen (1980) determined a total mass of  $2000 M_{\odot}$  using the virial theorem, a mean individual mass of  $2 M_{\odot}$ , and a velocity dispersion of  $0.7 \text{ km} \cdot \text{s}^{-1}$  in each spatial direction.

Lee & Sung (1995) measured  $700 M_{\odot}$  using the luminosity distribution and a mass-luminosity relation.

Pinfield et al. (1998) fitting a King profile to the PSD of the cluster members obtained  $735 M_{\odot}$ .

Adams et al. (2001) counting individual masses of candidate members within  $5.5^{\circ}$  obtained a total mass of  $690 M_{\odot}$ .

Converse & Stahler (2008) found  $820 M_{\odot}$  after adding the individual masses of 1245 candidate members of Stauffer et al. (2007). To obtain these masses they transformed the  $K$  and  $I - K$  magnitude and colour into masses using the mass-luminosity relation given by the theoretical isochrone models of Baraffe et al. (1998). Later, Converse & Stahler (2010) redid their analysis and found the total mass to be  $870 \pm 35 M_{\odot}$ .

## 2.7 The Pleiades DANCe DR2

The Pleiades DANCe DR2 contains astrometric (stellar positions and proper motions) and photometric ( $ugrizYJHK_s$ ) measurements for 1,972,245 objects. As explained in Chapter 1, the DANCe data set has an heterogenous origin as can be seen from Fig. 2.12. The interested reader can find more the details of this data set and its processing on Bouy et al. (2013). Here I briefly summarise its properties. Table 2.2 contains the basic statistics for the observables, while Table 2.3 does it for the uncertainties. As an example, Fig. 2.13 shows the proper motions uncertainties as a function of the  $i$  magnitude.

### 2.7.1 Selection of observables

As mentioned earlier, the Pleiades DANCe DR2 contains the positions  $R.A., Dec.$ , proper motions,  $\mu_{R.A.}, \mu_{Dec.}$ , and photometric  $ugrizYJHK_s$  bands, of almost 2 million sources on the vicinity of the Pleiades clusters. Although these 13 observables carry

**KPNO/Mosaic1**    **UKIRT/WFCAM**    **Subaru/SuprimeCam**    **CFHT/CFHT12K**  
**INT/WFC**    **CFHT/UH8K**    **KPNO/NEWFIRM**    **CTIO/MOSAIC2**    **CFHT/MegaCam**

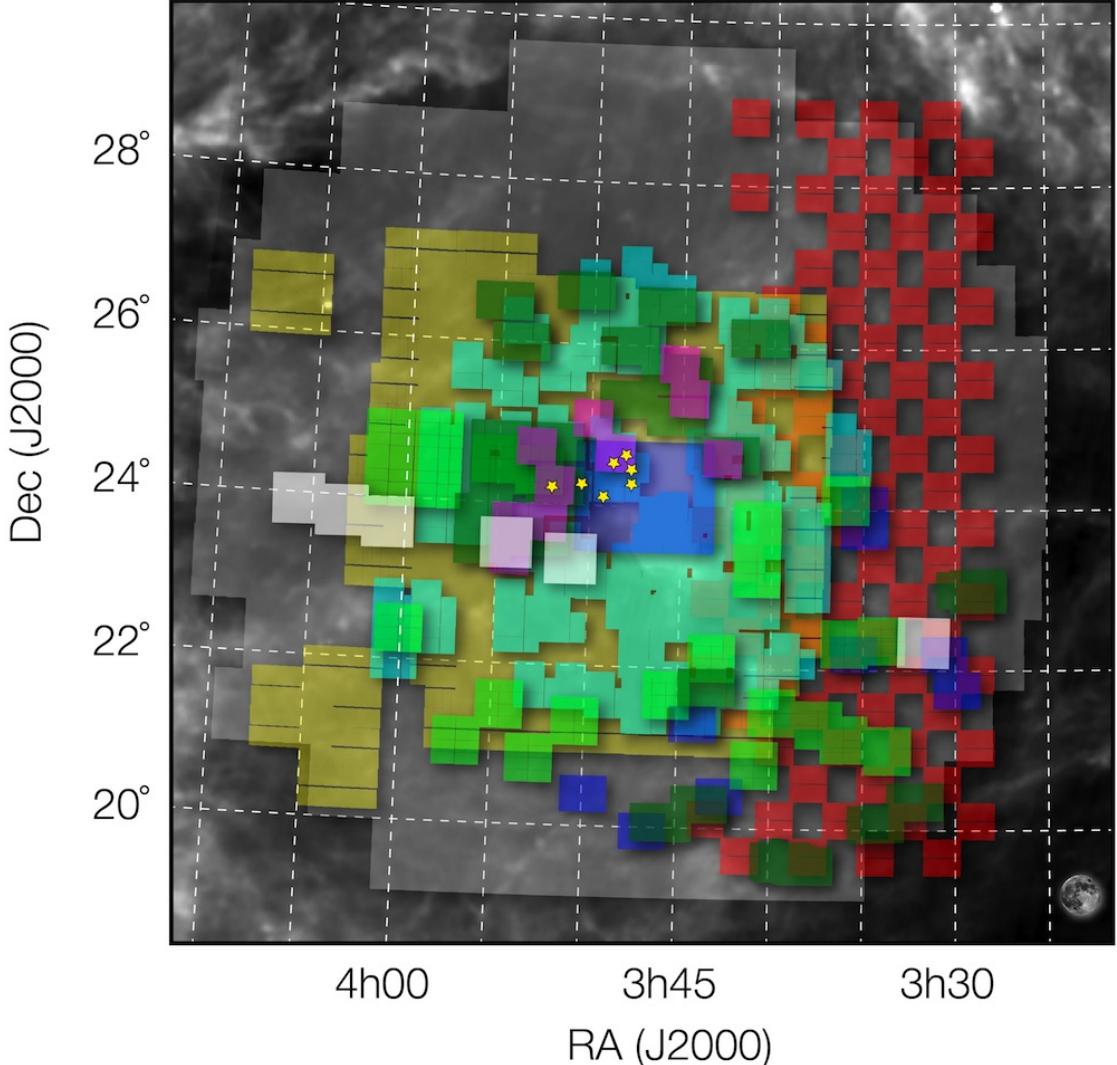


Figure 2.12: Patchy composition of the Pleiades DANCe DR2. The moon shows the scale, and the yellow stars correspond to the central brightest objects of the Pleiades cluster. Reproduced from Figure 1 of Bouy et al. (2013), *Dynamical analysis of nearby clusters. Automated astrometry from the ground: precision proper motions over a wide field*, A&A, Vol. 554.

information valuable to discriminate cluster members from field objects, not all of them discriminate in the same amount. Sarro et al. (2014) used random forest to select the observables that were the most discriminant. In their analysis, these authors did not include the stellar positions. They obtain that the AM, RF-2, and RF-3 reference sets containing the observables  $\mu_{R.A.}, \mu_{Dec.}$  proper motions, and photometric bands  $rizYJHK_s$  were the most discriminants. However, most of the objects with a missing  $r$  band occur at the faint end of the cluster sequence. Therefore, these authors trained their model in two stages. Their first stage used all literature candidate members

Table 2.2: Summary of the Pleiades DANCe DR2.

Observable	Min.	1st. Qu.	Median	Mean	3rd. Qu.	Max.	NA's
RA [deg]	51.23	55.40	57.35	57.26	59.01	62.94	0
Dec. [deg]	19.12	22.47	24.32	24.27	25.95	29.69	0
$\mu_\alpha [mas \cdot yr^{-1}]$	-99.998	-6.060	-1.645	-1.240	3.401	99.996	0
$\mu_\delta [mas \cdot yr^{-1}]$	-99.997	-2.835	2.548	1.976	7.017	99.989	0
u [mag]	13.6	20.4	22.0	21.6	23.3	25.2	1756374
g [mag]	9.4	19.6	22.1	21.1	23.3	25.5	1492564
r [mag]	8.4	17.6	21.3	20.3	22.6	25.1	1222853
i [mag]	7.5	20.0	21.6	21.0	22.7	25.5	820861
z [mag]	11.2	17.9	19.3	18.9	20.2	25.0	697412
Y [mag]	8.3	17.2	18.5	18.1	19.4	24.2	688144
J [mag]	2.8	16.7	17.9	17.5	18.8	23.1	645469
H [mag]	2.0	16.1	17.3	16.9	18.1	20.9	653682
$K_s$ [mag]	1.8	16.0	17.0	16.7	17.7	23.8	561745

Table 2.3: Uncertainties of the Pleiades DANCe DR2.

Observable	Min.	1st. Qu.	Median	3rd. Qu.	Max.	Mean
RA [deg]	8.900e-08	9.270e-07	1.933e-06	4.037e-06	2.156e-02	3.173e-06
Dec. [deg]	8.900e-08	9.270e-07	1.932e-06	4.037e-06	2.156e-02	3.173e-06
$\mu_\alpha [mas \cdot yr^{-1}]$	2.01e-01	1.89e+00	4.35e+00	1.00e+01	1.49e+22	3.99e+16
$\mu_\delta [mas \cdot yr^{-1}]$	1.92e-01	1.89e+00	4.35e+00	1.00e+01	4.71e+09	1.42e+04
u [mag]	3.73e-04	8.07e-03	3.06e-02	8.56e-02	2.17e-01	5.48e-02
g [mag]	1.72e-01	1.02e-02	3.90e-02	7.90e-02	1.82e+00	5.34e-02
r [mag]	2.83e-04	1.54e-02	4.88e-02	1.04e-01	1.42e+00	6.34e-02
i [mag]	4.04e-04	9.03e-03	2.73e-02	5.85e-02	2.40e+00	4.37e-02
z [mag]	6.49e-04	5.62e-02	9.16e-02	1.85e-01	3.12e+00	1.34e-01
Y [mag]	3.00e-02	5.21e-02	6.50e-02	1.03e-01	9.01e+00	8.56e-02
J [mag]	1.60e-02	5.24e-02	6.66e-02	1.04e-01	8.89e+00	8.57e-02
H [mag]	1.40e-02	5.28e-02	7.04e-02	1.10e-01	1.00e+01	8.85e-02
$K_s$ [mag]	1.40e-02	5.75e-02	8.17e-02	1.32e-01	3.88e+01	1.04e-01

with the  $r$  band observed. Then, they discarded the  $r$  band observations and continue the training of their model with the objects in which the  $r$  band was missing. In a subsequent analysis using roughly the same methodology, [Bouy et al. \(2015\)](#) skipped the first training stage and worked only with the RF-2, which also excludes the  $z$  band.

In this work, I select the  $\alpha, \delta, \mu_\alpha, \mu_\delta, i, Y, J, H, K_s$  observables as my reference set. The set of proper motions and photometric bands correspond to those used by [Sarro et al. \(2014\)](#), and [Bouy et al. \(2015\)](#).

Independent analyses will be performed to obtain: i) the kinematic and

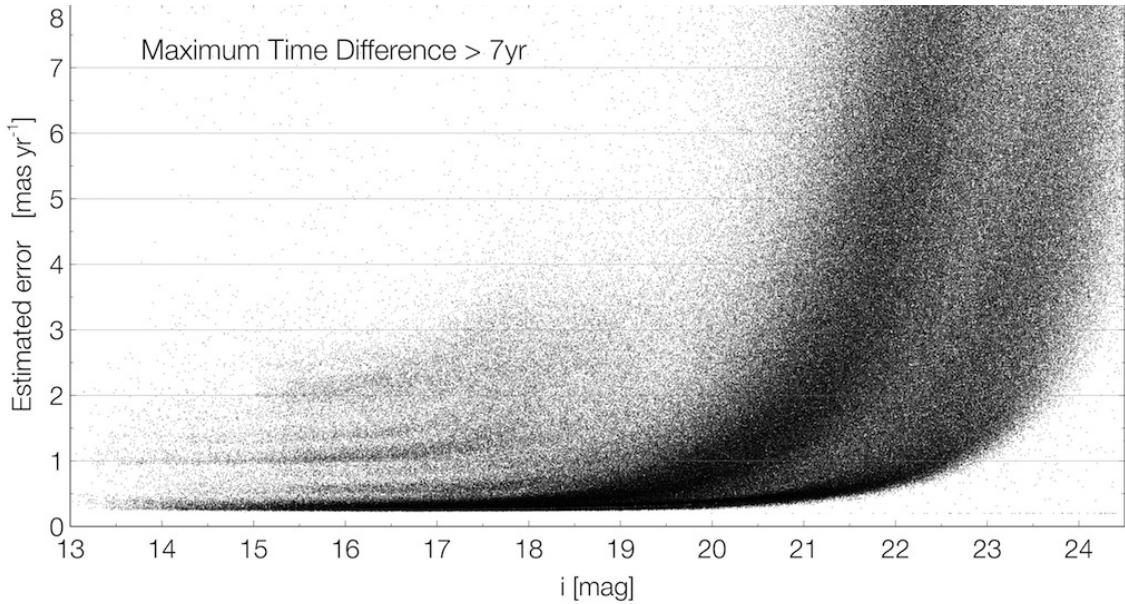


Figure 2.13: Proper motion uncertainty as a function of the photometric magnitude in the  $i$  band. Reproduced from Figure 12 of [Bouy et al. \(2013\)](#), *Dynamical analysis of nearby clusters. Automated astrometry from the ground: precision proper motions over a wide field*, A&A, Vol. 554.

photometric distributions of the cluster population, and ii) the PSD. This decision roots in the following reasons.

First. To fulfil the objective of the DANCe project (see Section 1.3), the inventory of kinematic and photometric distributions of the NYC must be obtained with an homogenous methodology. The DANCe NYC targets (see Tables 1.1 and 1.2) do not share similar projected spatial density profiles; while Taurus, Ophiucus and the Trapezium are extended (see for example [Simon 1997](#)), the Pleiades is almost radially symmetric ([Raboud & Mermilliod 1998](#)). Thus, including the spatial information together with the proper motions and photometry would require completely different models for the PSD of different clusters. That would result in non-homogenous methodologies that would bias any comparison between the derived PDMD of these clusters.

Second. To validate the results of the present work, which will be conducted in the Pleiades cluster, we must compare them with similar results under the most similar conditions. Since [Bouy et al. \(2015\)](#); [Sarro et al. \(2014\)](#) do not include the stellar positions in their observables, I do that as well.

Third. Almost all previous analyses of the Pleiades PSD use King profile (see Section 2.3) without providing any further reason that its physical interpretability, which nevertheless remains a good one. Thus, we decided to perform a Bayesian model selection analysis to compare how well the common surface density profiles, included the King's one, reproduce the Pleiades PSD.

The results of this analysis are in preparation and will be soon submitted to the *A&A* journal.

As will be described in Section 3.3.4, the photometry is modelled by parametric series of cubic splines. I choose the *true* colour index  $i - K_s$  (in the following  $CI$ ) to be the parameter of these series. This colour allows the most one-to-one dependent-independent variable relations. This one-to-one relation can be seen in Figures 2.14 and 2.15, where I show the colour-magnitude diagram of  $K$  vs  $i - K_s$  and  $K$  vs colours  $Y - K_s$ ,  $J - K_s$ ,  $H - K_s$ , and  $Y - J$ , respectively. This one-to-one relation is crucial to avoid degeneracies. Without it, two magnitudes could be described by the same colour index. Therefore a monotonic relation would not be valid.

Thus, our photometric set of observables is made of  $i - K_s$ ,  $Y$ ,  $J$ ,  $H$  and  $K_s$ .

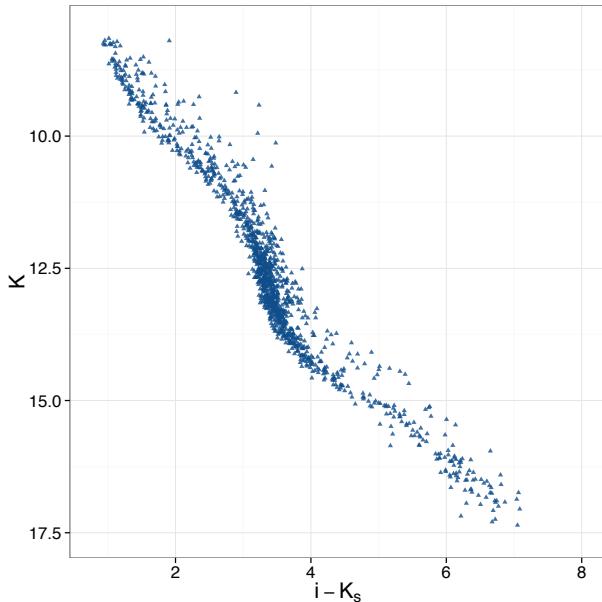


Figure 2.14:  $K_s$  vs  $i - K_s$  CMD for the Pleiades candidate members of Bouy et al. (2015)

## 2.7.2 Data preprocessing

Since both photometry and proper motions carry crucial information for the disentanglement of the cluster population, we restrict the data set to only those objects with observed proper motions, and also at least two photometric entries in our photometric set ( $i - K_s, Y, J, H, K_s$ ). **Objects with only one photometric entry, although theoretically could be included in the data set, are left aside due to computational reasons. Their treatment requires selection statements to choose between the univariate and multivariate computational libraries that proved to be computationally expensive.**

The previous restrictions exclude 22 candidate members of Bouy et al. (2015), which only

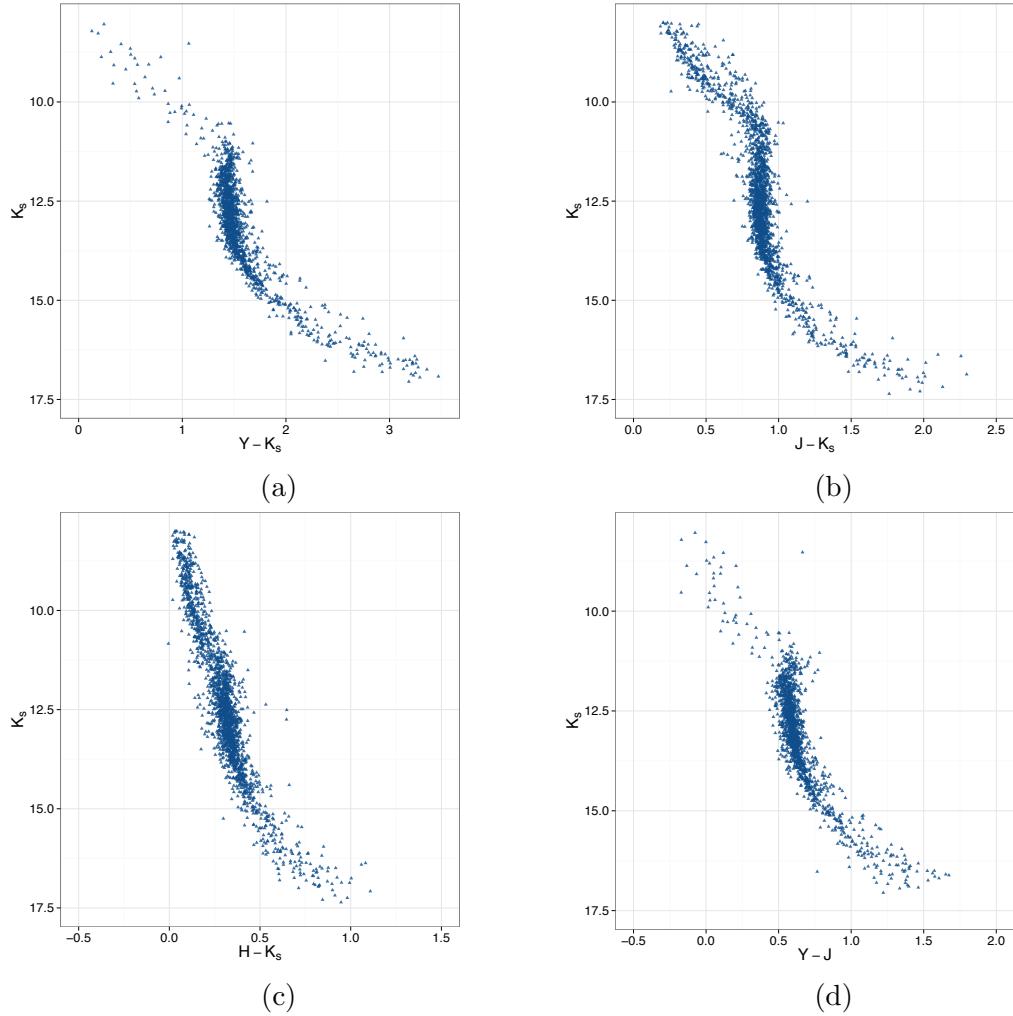


Figure 2.15: CMD for the Pleiades candidate members of [Bouy et al. \(2015\)](#). The magnitude  $K_s$  is shown versus the colour indices:  $Y - K_s$ (a),  $J - K_s$ (b),  $H - K_s$ (c), and  $Y - J$  (d).

have one observed value in the photometry. For these particular objects, we compute their marginal proper motion membership probability a posteriori, once the parameters of the model were inferred. The mode and 16 and 84 percentiles of their membership probabilities are listed in Table 2.4. Only four of these 22 objects have membership probabilities below 0.5, which may indicate they are contaminants. Since these membership probabilities were computed using only the kinematic information, these four members could not be discarded as probable candidate members.

Furthermore, we restrict the lower limit of the  $CI$  to the value of the brightest cluster member,  $CI = 0.8$  in the Pleiades DANCe DR2 data set. We do not expect to find new bluer members in the bright part of the CMDs. In the Tycho+DANCe data set ([Bouy et al. 2015](#)), which comprises the bright side of the cluster sequence, the bluer candidate member of [Bouy et al. \(2015\)](#) (for this data set) has a  $CI = 0.67$ . Showing that the cut at 0.8 is reasonable for the fainter DANCe data set. Also, we set the upper limit of the CI one magnitude redder than the colour index of the reddest known cluster member,

Table 2.4: Membership probabilities of the 22 excluded candidate members of [Bouy et al. \(2015\)](#).

ID DANCe	$P_{16}$	Mode	$P_{84}$
J035106.55+211604.3	0.751182	0.7774335	0.785560
J035057.42+240630.8	0.792295	0.8090186	0.829541
J034704.76+252249.8	0.701193	0.7319624	0.750745
J034725.80+250832.7	0.789872	0.8169538	0.827954
J034437.44+250815.6	0.762013	0.7773114	0.798125
J035125.88+244738.6	0.883488	0.9007972	0.905823
J034235.64+215029.7	0.838538	0.8662402	0.871439
J034516.66+243432.1	0.862284	0.8666611	0.881004
J034926.12+235714.8	0.852537	0.8606685	0.866286
J034920.60+244635.9	0.923319	0.9270399	0.935532
J035300.63+233252.3	0.762996	0.7747163	0.781901
J034606.52+235020.2	0.928688	0.9333306	0.940772
J035040.89+245657.7	0.509435	0.5215143	0.530379
J034845.33+233124.8	0.260551	0.2650812	0.275513
J034713.67+234953.3	0.814689	0.8489593	0.855902
J034546.48+234743.0	0.897035	0.9098442	0.912059
J034548.95+235110.2	0.933892	0.9376429	0.945558
J035202.26+242148.1	0.248011	0.2649874	0.305949
J035313.22+235540.8	0.518388	0.5425345	0.553376
J034425.60+244052.5	0.581242	0.5902633	0.602096
J035518.38+245637.2	0.198074	0.2087989	0.252831
J035418.93+252944.0	0.366009	0.3760922	0.386213

$CI = 8$ , thus allowing for new discoveries. Due to the sensitivity limits of the DR2 survey in  $i$  and  $K_s$  bands, ( $i \sim 23$  mag and  $K_s \sim 18$  mag, see Appendix A of Bouy et al. 2015), objects with  $CI > 8$  have  $K_s$  magnitudes brighter than 16 mag. This combination of CI and  $K_s$  magnitude is incompatible with the cluster sequence of Bouy et al. (2015) (see Fig. 2.16). Thus, we discard a priori these 262 objects as cluster members. Although formally these two cuts in the  $CI$  are not needed for the statistical analysis, they nevertheless improved significantly the computing time required for it. If we were to include these objects, the resulting  $CI$  range ( $CI \in [-6, 12.5]$ ) would have been 2.5 times larger than the  $CI \in [0.8, 8]$ . Thus, increasing in the same amount the computing time of the analysis (for more details, see Footnote 12 on page 69).

With the previous restrictions to the DANCe DR2 reduced it to 1 424 893 objects, with the largest number of rejected objects ( $\sim 500\,000$ ) resulting from a completely missing photometry.

Our computational constraints and the costly computations of our methodology (see Sect. 3.2), prevented its application to the entire data set. However, since the precision of our methodology, as that of any statistical analysis, increases with the number of independent observations, we find that a size of  $10^5$  source for our data is a reasonable compromise. Although a smaller data set produces faster results, it also renders a less precise and potentially more biased model of the field (in the area around the cluster) and therefore, a more contaminated model of the cluster. Thus, our data set was restricted to the  $10^5$  objects with highest membership probabilities according to Bouy et al. (2015). In the following I refer to this data set as the restricted DR2 or RDR2. The majority of these objects ( $\sim 98\%$ ) belong to the field with cluster membership probabilities about zero according to Sarro et al. (2014); Bouy et al. (2015). Thus, under the assumption that the membership probabilities given by these authors are correct, the probability of leaving out a cluster member is negligible. **The RDR2 data set is used to construct the cluster model and obtain the membership probability of its  $10^5$  objects.** For the remainder of the objects in the Pleiades DANCe DR2, we assign membership probabilities *a posteriori*, once the cluster model is constructed.

**The data set for the analysis of the Pleiades PSD consists of the high membership probability objects recovered after applying the BHM methodology (see Section 3.3) to the Pleiades DANCe RDR2 data set.** For the sake of simplicity, I call this data set the High Membership Probability set (HMPs). Although the Pleiades DANCe DR2 extends up to a radius of  $6^\circ$ , the spatial coverage is not homogeneous. As Fig. 2.12 shows, the deepest images of *UKIDS* cover the entire region while the rest of the instruments cover mainly the central part. Due to this patchy pattern, Bouy et al. (2015) used the central  $3^\circ$  region as the most complete one. Thus, for the the analysis of the PSD we define two samples, the complete one comprising those objects in the

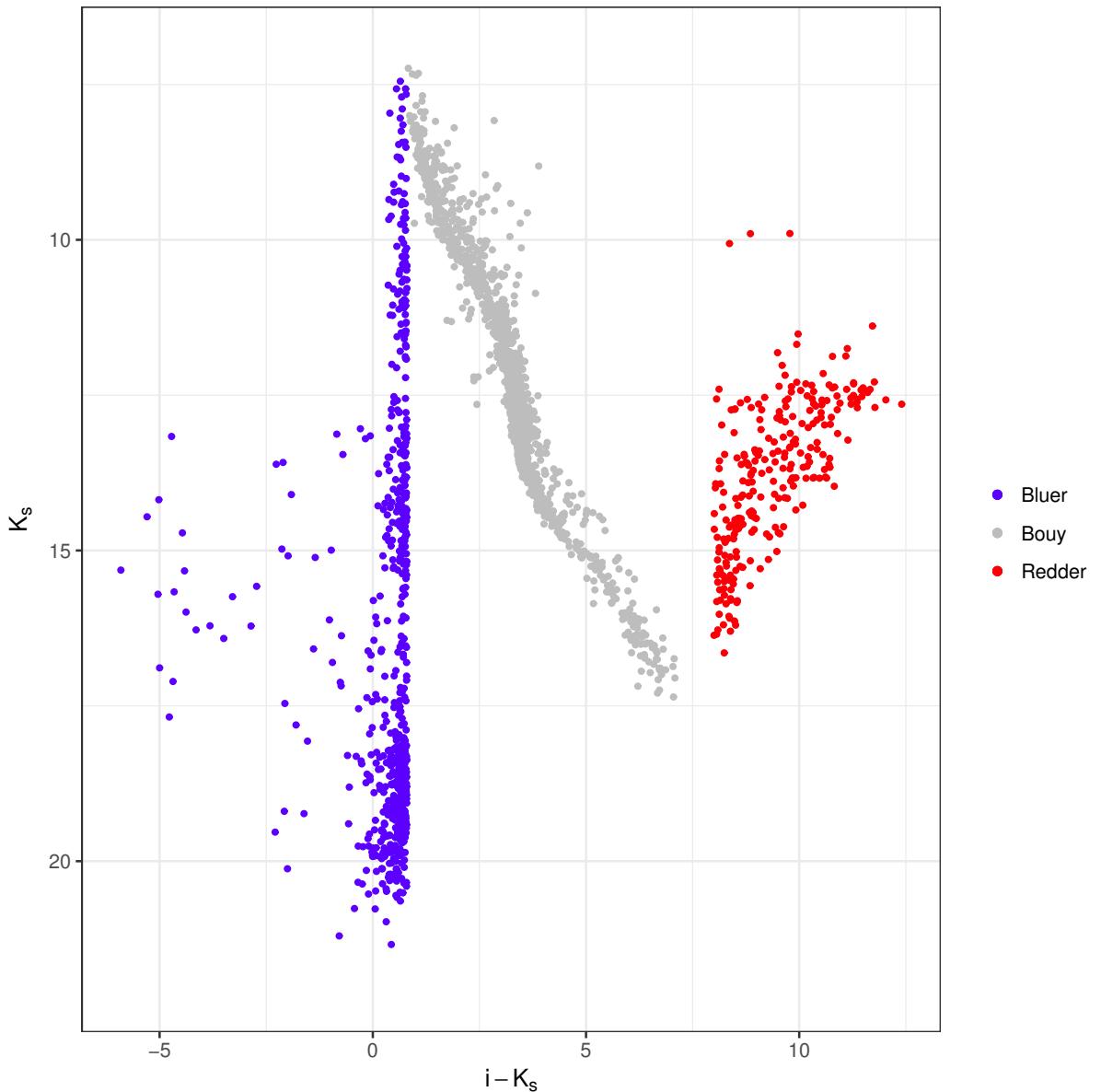


Figure 2.16:  $K_s$  vs  $CI$  CMD showing the candidate members of [Bouy et al. \(2015\)](#)(grey dots) and the DANCe DR2 objects bluer (blue dots) and redder (red dots) than the  $CI \in [0.8, 8]$ .

central  $3^\circ$ , and the extended one, which includes all objects up to a radius of  $6^\circ$ .

# Chapter 3

## Bayesian formalism

This chapter provides a general introduction to probability theory and its application to parametric inference. The objective of this work is to infer the probability distributions of the cluster properties (e.g. luminosity and velocity). Bayes' theorem provides the proper probabilistic framework for the inference of the parameters governing these distributions. The Bayesian framework demands, though, the setting up of prior beliefs about the parameters values. Thus, later in this chapter, I describe the reason why the Bayesian Hierarchical Models are the least subjective to establish priors. Once the posterior distribution of the parameters in the model has been analytically described, I proceed to describe the MCMC techniques and the particular one I use to sample the posterior distribution.

In the Sections ahead I also provide the details on the assumptions I make to model the data, and to choose the parameters of the prior distributions. The two final sections focus on the practical issues related to the sampling of the posterior distributions, and the description of the codes I adopted and/or developed.

Partial results of the work presented here have been submitted to the journal A&A as [Olivares et al. \(2017\)](#). In the following, I use both pronouns *we* and *I* to refer the investigation done by collaborators of the DANCe team (see Chapter 1).

### 3.1 Introduction to probability theory.

Uncertainty and probability are closely entangled. Every measurement has an associated uncertainty, otherwise is not a complete measurement<sup>1</sup>. The term uncertainty must not be confused with the term error, which refers to the difference between the measured value of the quantity and its *true* value<sup>2</sup> ([JCGM 2008](#)). It is commonly agreed that the uncertainty of a measurement can be expressed in a probabilistic basis ([JCGM 2008](#)). It

---

<sup>1</sup>Upper and lower limits are examples of incomplete measurements.

<sup>2</sup>The true value is that which ideally results when the uncertainty tends to zero.

means that whenever we measure a quantity,  $a$  for example, then the distribution of the repeated measurements of  $a$ , follows a probability distribution function,  $p(a)$ . Formally, if  $a$  is a discrete variable, then its probability distribution is called probability mass function (PMF). On the other hand, if  $a$  is continuous, then its probability distribution is called probability density function (PDF). Throughout the text, I refer to the probability distribution function of a random variable as its probability distribution or simply its distribution.

Any probability distribution satisfies the following properties:

**Property 1** It has units, those of the inverse of  $a$ .

**Property 2**  $p(a) \geq 0 \quad \forall a \in S_a$ , with  $S_a$  the support of  $a$ .

**Property 3**  $1 = \int_{S_a} p(a)da$ .

If  $a$  is a discrete variable, then the integral, in the last property, change to the sum of all possible values of  $p(a)$ .

These properties hold regardless of the dimension of  $a$ . Furthermore, they also hold for conditional probability distributions. A conditional probability distribution results from the knowledge about the particular value of one or several variables of the probability distribution. For example, be  $p(\alpha, \delta, \tau)$  the joint probability distribution of sky positions  $\alpha, \delta$  and time  $\tau$  of an object. Then, at the particular moment  $\tau = \tau_0$  (with  $\tau_0 \in S_\tau$ ) the object will have a probability distribution for its sky positions given by the conditional probability distribution  $p(\alpha, \delta | \tau_0)$ . Since,  $p(\alpha, \delta | \tau_0)$  is still a probability distribution on  $\alpha$  and  $\delta$ , it must also satisfy:

- It has units of  $\alpha^{-1}\delta^{-1}$ .
- $p(\alpha, \delta | \tau_0) \geq 0 \quad \forall \alpha \in S_\alpha, \delta \in S_\delta$ .
- $1 = \int_{S_\alpha} \int_{S_\delta} p(\alpha, \delta | \tau_0) d\alpha \cdot d\delta$ .

The link between joint and conditioned probabilities is given by the following symmetric definition:

$$\begin{aligned} p(a, b) &= p(a|b) \cdot p(b). \\ p(a, b) &= p(b|a) \cdot p(a), \end{aligned} \tag{3.1}$$

which can be further conditioned on  $c$  to obtain:

$$\begin{aligned} p(a, b|c) &= p(a|b, c) \cdot p(b|c), \\ p(a, b|c) &= p(b|a, c) \cdot p(a|c). \end{aligned} \tag{3.2}$$

If the joint probability of  $a$  and  $b$  can be factorised, this is

$$p(a, b) = p(a) \cdot p(b), \quad (3.3)$$

then  $a$  and  $b$  are say to be *independent*. An alternative option is to say that  $a$  and  $b$  are *independent* if the conditional probability of  $a$  on  $b$  is  $p(a|b) = p(a)$ .

**Property 3** establishes that the amount of probability density<sup>3</sup> (or mass if  $a$  is discrete) spread over the volume of the support, adds to one, thus keeping the integrals of PDFs bounded. Two important operations using these bounded integrals are the following.

The *marginalisation* of *nuisance* parameters. These are parameters that, although are necessary in the model, lack interest for the research. The classical example of a nuisance parameters is the standard deviation,  $\sigma$ , of a normal distribution when the interest lies solely in the mean,  $\mu$ . This nuisance parameter can be marginalised from the joint PDF of the parameters,  $p(\mu, \sigma)$ , in the following way,

$$p(\mu) = \int_0^\infty p(\mu, \sigma) \cdot d\sigma. \quad (3.4)$$

The computing of *expected values*. The expected value of  $a$ ,  $E(a)$ , corresponds to the mean of  $a$  once we have drawn many realisations from its probability distribution. To compute it, we add all the possible values of  $a$  weighted by their probability. This is,

$$E(a) = \int_a^\infty a \cdot p(a) \cdot da. \quad (3.5)$$

Once again, these last two equations (3.4 and 3.5) hold if the distributions are conditioned on any other measurement.

It is important to recall that the term measurement, and its unavoidable uncertainty, refer not just to directly measured quantities, like the photons (counts) and pixels in a CCD, but also to indirect measurements. Stellar magnitudes and positions in the sky, for example, are indirect measurements derived from the direct measurement of photons, pixels and telescope arrangements. This generalisation also applies to the measurement of parameters in any physical or statistical model.

This Section ends with a brief description of the procedure that is generally applied when we want to transform a probability distribution into another probability distribution, under a nonlinear transformation (for more details see for example [Bishop 2006](#), pages 18 and 19).

Let  $f$  be a probability distribution on  $x \subset \mathbb{R}$ , with support on  $a < x < b$ , then

---

<sup>3</sup>Which could be infinite, like in Dirac's delta.

$$\int_a^b f(x)dx = 1$$

Let  $y = g(x)$  be the nonlinear transformation, with  $g$  a function of  $x$  with inverse  $g^{-1}$ , continuous and with continuous derivative, so that  $x = g^{-1}(y)$ , with  $y \subset \mathbb{R}$ . Then, the following is true,

$$\int_a^b f(x)dx = \int_{g(a)}^{g(b)} f(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right| dy. \quad (3.6)$$

### 3.1.1 Bayes theorem

The definition of conditioned probability (Eq. 3.2) leads to Bayes' theorem:

$$p(a|b, c) = \frac{p(b|a, c) \cdot p(a|c)}{p(b|c)}. \quad (3.7)$$

Integrating on  $a$  we find that,

$$\begin{aligned} p(b|c) \cdot \int_a p(a|b, c) \cdot da &= \int_a p(b|a, c) \cdot p(a|c) \cdot da \\ p(b|c) &= \int_a p(b|a, c) \cdot p(a|c) \cdot da. \end{aligned} \quad (3.8)$$

This Equation illustrates that  $p(b|c)$  is a normalisation constant which can be evaluated once  $p(b|a, c)$  and  $p(a|c)$  are known. This turns out to be very useful, since it tells us that  $p(a|b, c) \propto p(b|a, c) \cdot p(a|c)$ .

### Models and parametric inference

In a broad sense, models are representations or abstractions of the knowledge someone has about something. Sometimes this knowledge is also shared by others. Models are everywhere in our daily life: from the words we speak every day, to the evolution of the species and the general relativity; from a kid's drawing to cosmological models. In science, however, the concept of model is restricted to a mathematical representation of the relations (the knowledge) among the entities that the model attempts to describe: the observables (i.e. the data). If the model contains variables that through the different values they take reproduce in some extent the observables, then model is called parametric and the variables the parameters.

Parametric statistical models, like the ones I use in this work, assume that the underlying population of interest, from which the observed data is just a sample, can be described by parametric probability distribution functions. The act of finding the parameters governing these distributions is called parametric inference. This last can focus either on the entire PDFs of the pa-

rameters, or just on some summary of them (e.g. the maximum a posteriori, also known as MAP, or the mean and variance).

The proper way to obtain the entire probability distribution of the parameters in a model, given the data, is through Bayes' theorem. Thus it is called Bayesian inference. Another example of parametric inference is the maximum likelihood approach, where the likelihood, which is seen as a function of the parameters, is maximised. Despite that it obtains the parameter values that make the model to resemble the data, it does not return their probability distribution. Formally, the likelihood is a probability distribution for the data, and just a function of the parameters. Thus, to obtain the probability distribution of the parameters, the likelihood must be multiplied by the priors and the product normalised. This is what Bayes' theorem does.

In this context, Bayes' theorem is:

$$p(\boldsymbol{\theta}|\mathbf{D}, \mathcal{M}) = \frac{p(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}) \cdot p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathbf{D}|\mathcal{M})}. \quad (3.9)$$

where  $\boldsymbol{\theta}$ ,  $\mathbf{D}$  and  $\mathcal{M}$  correspond, respectively, to the parameters in the model, the data which the model tries to describe, and the model itself.

The term on the **left-hand side** is called the posterior probability distribution of the parameters,  $\boldsymbol{\theta}$  given the data  $\mathbf{D}$ , and the model  $\mathcal{M}$ . On the right hand side, the two terms in the numerator are called the *likelihood* for the data,  $p(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M})$  and the *prior* of the parameters,  $p(\boldsymbol{\theta}|\mathcal{M})$ . The denominator,  $p(\mathbf{D}|\mathcal{M})$ , is called the *evidence*, and results from

$$Z \equiv p(\mathbf{D}|\mathcal{M}) = \int_{\boldsymbol{\theta}} p(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}) \cdot p(\boldsymbol{\theta}|\mathcal{M}) \cdot d\boldsymbol{\theta} \quad (3.10)$$

Formally, the likelihood and the prior are probability distributions for the data  $\mathbf{D}$  and of parameters  $\boldsymbol{\theta}$ , respectively. However, for the posterior to be a probability distribution of the parameters, it only suffices that the product of the likelihood times the prior does not vanish everywhere or be negative anywhere<sup>4</sup>. If these are not probability distributions, they are called *improper* priors or *improper* likelihoods. In the extreme case that their product vanishes everywhere, which may be the case if the prior is terribly specified or if the likelihood does not take proper account of extreme data, the posterior will not be a probability distribution due to a division by zero. Nevertheless, it makes no sense to try to estimate the parameters of a model with zero evidence.

As mentioned before, the likelihood is a probability distribution for the data, given the parameters, regardless of the size of it. Almost always the data is a collection of measurements of several objects, but it could also be made up of just one object. The collection of measurements of one or several quantities of a single object follows a probability distribution, which is usually summarised by two statistics. It is often assumed that this

---

<sup>4</sup>See Property 2. Although negative probabilities may have sense in quantum mechanics. See for example Dirac (1942)

probability distribution is normal (univariate or multivariate), which then is summarised by the mean and the standard deviation. These two are commonly known as the datum and its uncertainty, respectively. A data set is then composed of the collection of summary statistics of one or several objects. To compute the likelihood for this collection of statistics (i.e. the data), some assumption must be made.

When measuring the properties of objects, it is often assumed that the probability distribution obtained from the collection of measurements of a single object, is independent from that of another object. For example, if we were to measure the weight of a group of persons, we usually assume that the PDF of the weight of one person, is independent of that of another person. It means that measuring the weight of one person has no effect at all in the weight of another person.

In this work it is always assumed that the PDFs of the measured quantities of individual objects are independent amongst them. Nevertheless, in the following, I give an example where this assumption may not be entirely right. Obtaining stellar positions in celestial coordinates often requires what is called an astrometric solution. This solution is a map from the raw data, like pixel positions in the detectors and observing epoch, to the celestial coordinates (e.g. right ascension, declination and epoch). This astrometric solution often needs large collections of measurements of the same objects, so that they can be robustly estimated. Since this mapping is computed from the data (e.g. using maximum-likelihood estimates) and then applied to the same data, then it is common to observe correlations among the uncertainties of different objects (see for example [Holl et al. 2010](#); [Gaia Collaboration et al. 2017](#)). If the correlation in the uncertainties is significative, then their PDFs are probably not independent<sup>5</sup>.

Let  $\mathbf{D}$  be the data, i.e. the collection of statistics of the  $N$  objects, and  $p_n(\mathbf{d}_n)$  be the probability distribution rendered by several measurements of object  $n$ . If these  $\{p_n(\mathbf{d}_n)\}_{n=1}^N$  are assumed to be independent, then

$$p(\mathbf{D}) = \prod_{n=1}^N p_n(\mathbf{d}_n), \quad (3.11)$$

with  $p_n$  explicitly stating that the individual probability distributions are distinct.

Similarly, if the likelihood of the data,  $p(\mathbf{D}|\boldsymbol{\theta}, M)$  is assumed to be independent for each object, then

$$p(\mathbf{D}|\boldsymbol{\theta}, M) = \prod_{n=1}^N p(\mathbf{d}_n|\boldsymbol{\theta}, M). \quad (3.12)$$

The term  $p(\mathbf{d}_n|\boldsymbol{\theta}, M)$  is the likelihood of datum  $\mathbf{d}_n$ . This is also called the *generative* model, since it contains the necessary information to generate the data.

---

<sup>5</sup>Independent PDFs produce uncorrelated samples, however, uncorrelated samples do not imply independence between the PDFs of their underlying populations.

To take into account the uncertainty process for object  $n$ , we model the datum  $\mathbf{d}_n$  as resulting from the addition of the true value,  $\mathbf{x}_n$ , which is given by the model, with a random variable,  $\mathbf{e}_n$ , given by the uncertainty process. This is

$$\mathbf{d}_n = \mathbf{x}_n + \mathbf{e}_n.$$

In general, if the model does not contain any intrinsic dispersion, then its likelihood can be thought of as a Dirac  $\delta$  function, which is centred at the *true* value  $\mathbf{x}_n$ . Then, it can be added (i.e. convolved<sup>6</sup>) to the distribution of the uncertainty process. However, it is customary to also include an intrinsic dispersion in the model to account for over-simplistic assumptions or underestimated uncertainties.

In particular, it can be assumed that both the uncertainty process of datum  $\mathbf{d}_n$ , and the likelihood of the model are normally distributed with variances  $\Sigma_n^2$  and  $\Sigma^2$ , respectively,

$$\begin{aligned} p(\mathbf{e}_n | \Sigma_n^2) &= \mathcal{N}(\mathbf{e}_n | 0, \Sigma_n^2), \\ p(x_n | \boldsymbol{\theta}, M, \Sigma^2) &= \mathcal{N}(\mathbf{x}_n | \boldsymbol{\theta}, M, \Sigma^2). \end{aligned}$$

Formally,  $\Sigma^2$  is part of the set of model parameters,  $\boldsymbol{\theta}$ , but I explicitly leave it outside to exemplify the process.

Then the addition of these two normally distributed random variables results in another normally distributed random variable<sup>7</sup>. Thus,

$$p(\mathbf{x}_n | \boldsymbol{\theta}, M, \Sigma^2) * p(\mathbf{e}_n | \Sigma_n^2) = \mathcal{N}(\mathbf{d}_n | \boldsymbol{\theta}, M, \Sigma_n^2 + \Sigma^2). \quad (3.13)$$

Therefore, Eq. 3.12, can be expressed in general as,

$$p(\mathbf{D} | \boldsymbol{\theta}, M) = \prod_{n=1}^N p(\mathbf{d}_n | \boldsymbol{\theta}, M, \mathbf{u}_n), \quad (3.14)$$

where  $\mathbf{u}_n$  is the uncertainty of datum  $\mathbf{d}_n$ .

Bayes' theorem can be interpreted as the probabilistic way to update the state of knowledge. To me, it embodies the process of knowledge improvement once we recognise that knowledge is uncertain. Even when its uncertainty is negligible under the evidence that supports it. Bayes' theorem helps us update our prior beliefs once we multiply it by the likelihood of the data. Then, the posterior probabilities, become our new state of knowledge.

Furthermore, Bayes' theorem also provides the objective way to compare two models or hypothesis, and update the a priori knowledge used to construct them. This is called model selection, which I briefly explain in the next section.

---

<sup>6</sup>The addition of two random variables results in another random variable. This is analogous to the convolution, denoted  $*$ , of their PDFs.

<sup>7</sup>The convolution of two Gaussian PDFs is another Gaussian PDF.

### 3.1.2 Model Selection

Whenever we have a data set and two or more models that attempt to describe these data, the most straightforward thing to do is to compare these models. Almost always, we want to select the *best* model. Obviously the term *best* depends on the objective of the research. For example, imagine that our data set consists of the positions of an object as function of time. If we were interested in reproducing exactly the same points in the data set, the *best* model would be a polynomial with degree equal to the number of points. This polynomial will pass through all the points. However, once we recognise the unavoidable uncertainty of the data, we realise that an exact representation of the data may be of no use since it fits also the noise.

In general, we are interested in the predictive capabilities of a model, its ability to predict future observations rather than to replicate the ones we currently have. Thus, an exact representation of the observed data (an over-fitted model as in the previous example), will poorly describe any new data. In this sense, an over-fitted model *memorises* the data rather than *learns* from it.

A model that *learns* from the data is that which recovers the *true* underlying relation embedded in the data. This *true* underlying relation is the one that produces the *true* data. The observed data results once the uncertainty is added.

Nevertheless, we still need to select among different learning models.

We can draw some help from the commonly known Ockham's razor or principle<sup>8</sup>. It says:

*Among competing hypotheses, the one with the fewest assumptions should be selected.*

Here, hypotheses can be identified with models. Thus, this principle tells us we should choose the model that makes the fewest assumptions. I classify the assumptions of a model in two groups: fixed and free ones. The fixed assumptions belong to what I previously described as the *a priori* knowledge used to construct the model. These may render the model more interpretable in the physical or statistical sense, or even give it coherence within the corpus of a theory. The free assumptions on the other hand, correspond directly to the parameters in the model. They give it flexibility when fitting the data<sup>9</sup>. For example, in the case of a straight line model, the fact that the data is linearly related can be considered as a fixed assumption. The free assumptions correspond to the slope and ordinate at the origin.

When comparing a linear model to a quadratic one in which the constant term has been fixed, we see that they have the same number of free parameters, two, but clearly the

<sup>8</sup>The origin of this motto and its exact phrasing is beyond the scope of this work. I just mention that paradoxically, an ancient formulation is attributed to Ptolomey: "We consider it a good principle to explain the phenomena by the simplest hypothesis possible" ([Franklin 2002](#))

<sup>9</sup>However, they can also introduce degeneracy in the parametric space.

second one has an extra fixed assumption. Therefore, choosing the model with fewer free parameters does not necessarily mean choosing the model with the fewest assumptions.

One of the great advantages of the Bayesian methodology is that it incorporates directly Ockham's principle. Suppose that we want to compare two models,  $M_1$  and  $M_2$ , which we assume describe the data set,  $\mathbf{D}$ . Each model has prior probabilities,  $p(M_k)$  and likelihoods  $p(\mathbf{D}|M_k)$  (with  $k = 1, 2$ ). Notice that now, I use Bayes' theorem for models and not for parameters within a model. So, the prior probabilities of the models reflect our beliefs about the fixed assumptions within each model. On the other hand, the likelihood of the data, given the model, is related to the parameters (the free assumptions) and their prior probabilities, both within a model. This likelihood of the data given the model corresponds to the *evidence* of the model (Eq. 3.10). This evidence, written in terms of the model parameters,  $\theta_k$ , is now

$$p(\mathbf{D}|M_k) = \int_{\theta_k} p(\mathbf{D}|\theta_k, M_k) \cdot p(\theta_k|M_k) \cdot d\theta_k. \quad (3.15)$$

Bayes' theorem applied to models, instead of individual parameters as illustrated above, tells us that

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) \cdot p(M_k)}{p(\mathbf{D})}, \quad (3.16)$$

with  $k = 1, 2$ . Since there are only two models, their prior probabilities are related by  $p(M_1) = 1 - p(M_2)$ . Therefore,

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) \cdot p(M_k)}{p(\mathbf{D}|M_1) \cdot p(M_1) + p(\mathbf{D}|M_2) \cdot p(M_2)}. \quad (3.17)$$

From this last Equation, the ratio of the posterior distributions is:

$$\frac{p(M_1|\mathbf{D})}{p(M_2|\mathbf{D})} = \frac{p(\mathbf{D}|M_1) \cdot p(M_1)}{p(\mathbf{D}|M_2) \cdot p(M_2)}. \quad (3.18)$$

This ratio provides an objective measure of how better the model  $M_1$  is when compared to model  $M_2$ , under the measure provided by the data  $\mathbf{D}$  by means of the evidence. When both prior probabilities  $p(M_1)$  and  $p(M_2)$  are set alike, the ratio of posteriors equals the ratio of likelihoods. This is known as the *Bayes factor* (for a similar derivation and some examples of its application see [Kass & Raftery 1995](#)).

Even when the priors for the models are set alike, the evidences themselves (Eq. 3.15) embody Ockham's principle. The evidence is the integral, in parametric space, of the prior times the likelihood, with the likelihood acting as a weight to the priors. Then, the larger the number of parameters (free assumptions) is, the greater the volume over which the integral must be carried on, and the most spread the prior probability gets. Thus, unless the likelihood increases as well, the evidence is smaller in models with larger number of parameters.

As explained in this section, the paramount importance of Bayes' theorem comes from the fact that it is the proper probabilistic way to update knowledge based on the **statistical** evidence.

### 3.1.3 Membership probability

In the previous Section we **derived** the ratio of the probabilities of two competing models  $M_1$  and  $M_2$ , given the data  $\mathbf{D}$ . In this Section, I describe a similar problem: the probability of two competing models given the two likelihoods for a single datum,  $\mathbf{d}$ . It also can be interpreted as the probability that the datum  $\mathbf{d}$  was generated by model  $M_k$ . This probability is commonly known as the membership probability of the datum  $\mathbf{d}$  to belong to model or class,  $M_k$  ( $k = 1, 2$ ).

Bayes' theorem for this particular case is,

$$p(M_k|\mathbf{d}) = \frac{p(\mathbf{d}|M_k) \cdot p(M_k)}{\sum_{k=1}^2 p(\mathbf{d}|M_k) \cdot p(M_k)}, \quad (3.19)$$

where  $p(\mathbf{d}|M_k)$  is the likelihood of datum  $\mathbf{d}$  and,  $p(M_k)$  is the prior probability of model  $M_k$ .

## 3.2 Bayesian Hierarchical Models

### 3.2.1 Generalities

The Bayesian formalism requires the establishment of priors. These represent the beliefs the user of the model has about the possible values that parameters of the model can take **before new data are observed**. This is indeed subjective. This subjectivity is the main source of criticism from the non-Bayesian community <sup>10</sup>.

Bayesian hierarchical models, in the following BHM, are classified within the Empirical Bayes methods. On these methods, the prior distributions are inferred from the data rather than being directly specified, as it is done in common Bayesian methods. In BHM the priors are specified by parametric distributions whose parameters are also drawn from another parametric distribution in a hierarchical fashion. For this reason, hierarchical models are also called multilevel models. A full-BHM is that in which the parameters at the higher hierarchy **level** are drawn from a non-parametric distribution. In a non full-BHM the settlement of parameters stops at some level. These end-level parameters are called hyper-parameters.

Given their properties, BHMs represent the most objective way to the establishment of prior distributions (Gelman 2006). Regardless of the level of the BHM, for it to be effective, the family, or class, of prior distributions must be carefully chosen. These families

---

<sup>10</sup>See Gelman (2012) for a discussion on the ethical use of prior information

must allow the *true* value of the parameter of interest ([Morris 1983](#)). If the likelihood **parametric support** is not fully contained in the prior (for example when the likelihood, as a function of the parameters, has a maximum outside the domain of a truncated prior), then the inferred posterior **can** be biased. For this reason, inspecting the prior knowledge is an important step in any Bayesian study.

Despite their theoretical advantages, BHMs are difficult to evaluate since they require far more parameters than standard Bayesian methods. Furthermore, their hierarchy (levels) must stop at some point. There are at least two approaches to stop this hierarchy. The first one uses a non-parametric distribution for the parameters at the higher level. This renders, as previously noted, a full-BHM. **However, it demands an a priori knowledge of the non-parametric distribution, which, most of the time, is not the case.** The second option is to give a point estimate, usually the mean or the mode, for the distribution of the parameter at the top of the hierarchy.

Although in BHMs the parameters of the prior distributions are inferred from data, the user of the model has the important task of specifying the family of distributions to be used for them. Selecting these families continues to be an active area of research. The most common approaches are the following.

- Conjugate priors. In this kind of priors the posterior distribution, which results of the product of the prior times the likelihood, is also in the same family distribution of the prior.
- Default priors. These priors are used when there is insufficient information to set the prior and are supposed to be the default choice in the absence of any other information.
- Reference priors. Formally, a reference prior is permissible and maximise the missing information. See [Berger et al. \(2009\)](#) and references therein.
- Non-informative priors. This kind of priors intentionally discards all information on the phenomenon.
- Weakly informative priors. These provide intentionally weaker information than what actually is.

The default and non-informative approaches are discarded since for the Pleiades, which is one of the most studied cluster in the literature, there is already reliable information which must be used. Conjugate priors do not always agree with the previous knowledge. Reference priors, although an interesting approach, require complicated derivations which are hard to implement and thus computationally expensive.

From the previous options, I choose the weakly-informative approach. These are recommended by [Gelman \(2006\)](#); [Gelman et al. \(2008\)](#); [Huang & Wand \(2013\)](#) and

[Chung et al. \(2015\)](#), since they show better computational performance when compared to non-informative priors.

Whatever family used for the prior distribution, we must always analyse the priors in terms of the posterior distribution, and check if the late makes sense ([Gelman 2006](#); [Gelman et al. 2013](#), Chap. 6).

### 3.2.2 Examples

Since BHMs usually need more parameters than standard techniques, its use was restricted until fast computers were widely available. The concept of BHM was already present in the 1960s. However, it was not until the 1970s that they were used to infer parameters of normal distributions and linear models (see [Good 1980](#), for an historical perspective of BHMs). In modern days, BHMs have a wide range of applications. Just to cite some examples, [Gelman & Hill \(2007\)](#) use them in the social sciences, [Fei-Fei & Perona \(2005\)](#) applies them for vision recognition and, [Diard & Bessiere \(2008\)](#) for robot navigation.

BHMs are also widely applied in astrophysics. Although, originally its use was mainly in the domain of inference of cosmological parameters (see for example the works of [Feeney et al. 2013](#); [March et al. 2014](#); [Anderes et al. 2015](#); [Shariff et al. 2016](#); [Alsing et al. 2017](#)), they were rapidly adopted in other domains. For example, they have been used to study the eccentricity distribution of binary stars [Hogg et al. \(2010b\)](#), the Cepheids ([Barnes et al. 2004](#)) and RR Lyrae distances ([Jefferys et al. 2007](#)), the chemical composition ([Wolfgang & Lopez 2015](#)) and albedos of exoplanets ([Demory 2014](#)), extinction maps ([Sale 2012](#)), stellar parameters ([Shkedy et al. 2007](#)), and the present day mass distribution ([Tapiador et al. 2017](#)).

### 3.2.3 Graphical representation.

Probabilistic Graphical Models (PGM) provide the background to graphically depict BHMs. PGMs are graphs that portray the conditional relations among the elements in a probabilistic model. These elements can be constants or stochastic variables that interact by means of conditional relations, which can in turn be deterministic or stochastic.

In PGMs, stochastic variables are represented with circles and constants with squares. If the variable is known, as in the case of the data, it is represented with a filled symbol, otherwise with an empty symbol. Stochastic and deterministic relations are depicted with solid and dashed lines, respectively. If there is no line between two given elements, it indicates that they are assumed to be independent. Variables that repeat together, as in the case of the data, are grouped within a plate. The number of repetitions is indicated in one corner of the plate. For more details on PGMs see for example the book of [Koller & Friedman \(2009\)](#).

To exemplify the use of PGMs, Figure 3.1 shows the PGM for the BHM that infers the parameters in a Gaussian Mixture Model. In this model the parameters are the means  $\mu_k$ , variances  $\sigma_k^2$  and fractions  $\phi_k$  of each of the  $K$  gaussian distributions in the mixture. Then  $\mu_0$ ,  $\sigma_0^2$ ,  $\lambda$ ,  $\nu$  and  $\beta$  are the hyper-parameters, while  $x_i$  represent the data, and  $z_i$  the categorical latent variable that indicates the parent gaussian for datum  $x_i$ . The squiggly line ending in T indicates that  $z_i$  is switch, which selects the gaussian at which  $x_i$  belongs.

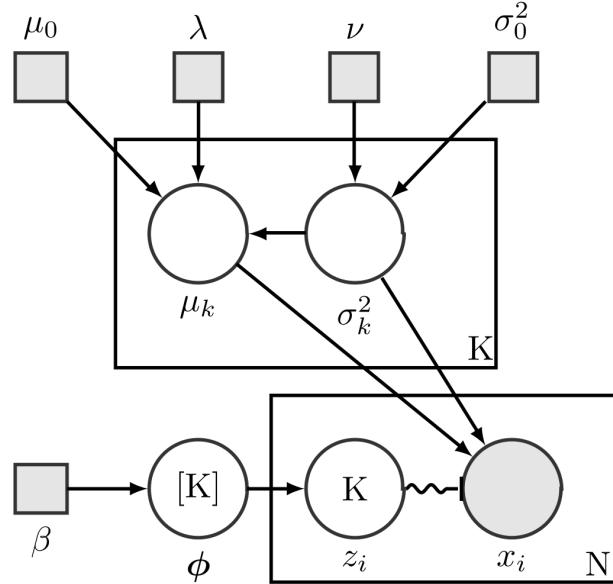


Figure 3.1: PGM representing the parametric inference of a Gaussian Mixture, see text for details. Figure by Benwing, license: Creative Commons BY-3.0

### 3.3 The BHM for the Pleiades DANCe data set

Creating a model is a complex task. As previously mentioned, a model is a mathematical representation of the knowledge about a certain phenomenon. Thus, a model demands the gathering and arranging of the *a priori* knowledge of the phenomenon as well as the gathering process of its associated data set.

The current understanding of the Pleiades cluster together with the description of the DANCe DR2 data set are summarised in Chapter 2.

Once the *a priori* knowledge has bee gathered, the creation of the model becomes an iterative and continuous process. Assembling the knowledge into a coherent system demands continuous iterations of decision making, test, and analysis of preliminary results. This Section provides a snapshot of this process: the state of the model once the article Olivares et al. (2017) was submitted.

Section 3.3.1 explains the statistical procedure to deal with one of the crucial aspects of the Pleiades DANCe DR2 data set: objects with missing values in their measurement

vector. Later, Section 3.3.2 provides details of how the relevant knowledge about the Pleiades cluster and field populations are embedded in the generative model of the data.

### 3.3.1 Missing values

Missing values refer to a non-measured or non-available (NA) value in the vector of measurements of an object. They can arise due to different statistical or physical processes.

From the physical perspective, they occur due to faint or bright sources that produce counts values outside the dynamical range of the detector. They can also emerge due to detector malfunctions, e.g. electronic failures, or to random effects, e.g. cosmic rays.

From the statistical perspective, important aspects of missing values are, i) the probability distribution of the sources in which they occur, and ii) if they originate due to censoring or truncation.

In the DANCe DR2, missing values occur only in the photometric measurements (see Table 2.2), with the bluer bands being the most affected. As expected, the probability distribution of sources with missing values is not **uniform**. Missing values occur with higher probability at the brightest and faintest ends of the photometric distributions ( $\sim 11$  mag and  $\sim 18$  mag in  $J$ ,  $H$  and  $K_s$  bands, respectively), see Figure 3.2. This is a crucial aspect for the statistical analysis. If missing values were **uniformly** distributed, then, any statistical analysis derived using only objects with completely observed vectors would be unbiased. This will occur because the objects with completely observed vectors would be a random sample of the data set. **Since this is not the case in the DANCe DR2, objects with missing values must be incorporated in the data set and properly treated to avoid biases.**

**Truncation occurs when the data set does not contain records, neither of the measured value nor of the number of measurements, for objects whose measured quantity lies outside certain limits.** On the other hand, censoring happens when the data set contains only partial information about the actual value of the measured quantity. Basically, censoring happens when the value lies outside the upper or lower limits of the measuring instrument, and these limits become the only available record of the measurement. Currently, the DANCe DR2 does not provide upper or lower limits for the censored data. Although these limits could also be inferred from the data, the heterogeneous origin of the DANCe DR2 prohibits this task. Furthermore, the missing values in the DANCe DR2 data set do not occur only because of censoring. Indeed, cosmic rays, hot pixels, halos of bright stars, diffraction patterns and cross-matching failures are just some examples of potential sources of missing values. The statistical treatment of missing values originating from all these sources lies beyond the scope of this work. However, there is a general way to deal with missing values **whatever** their origin.

In terms of probability, there is no distinction between missing values and parameters. Therefore, we can marginalise missing values as we do with any other nuisance parameter.

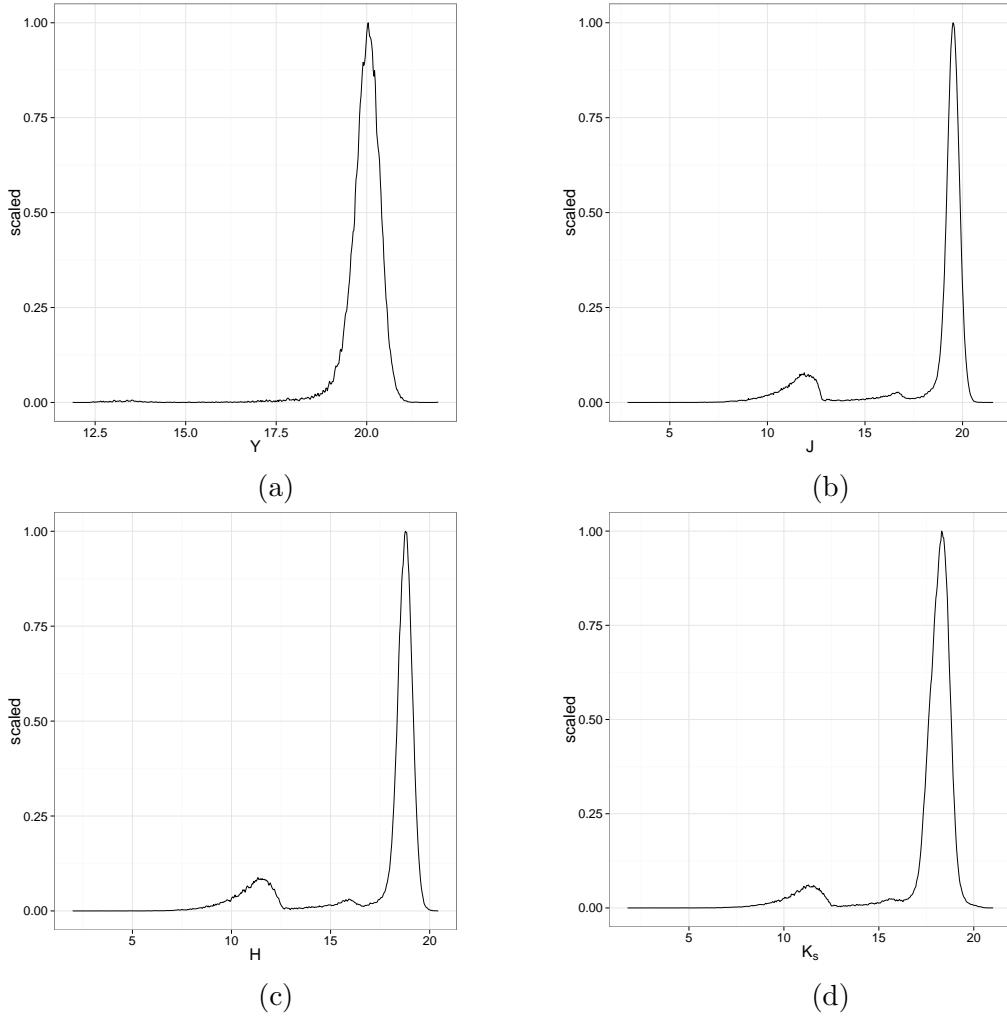


Figure 3.2: Probability distributions of objects with missing values. The majority of objects with missing values are located at the faint end of the magnitude distributions ( $K_s \sim 18$ ), while a non negligible fraction is at the bright end ( $K_s \sim 12$ ). The small bump at  $K_s \sim 16$  may have its origin in the complex pattern of spatial+photometric coverage, see [Bouy et al. \(2013\)](#).

If the vector of measurements  $\mathbf{d}$  has a missing value in entry  $i$ , then  $x_i = NA$ , and  $\mathbf{d} = \{x_1, x_2, \dots, x_{i-1}, NA, x_{i+1}, \dots, x_n\}$ , with  $n$  the dimension of  $\mathbf{d}$ . Thus, the marginalised likelihood of this datum, given model parameters  $\boldsymbol{\theta}$  is

$$p(\mathbf{d}|\boldsymbol{\theta}) = \int_{-\infty}^{\infty} p(\{x_1, x_2, \dots, x_{i-1}, x_i = NA, x_{i+1}, \dots, x_n\}|\boldsymbol{\theta}) \cdot dx_i. \quad (3.20)$$

Throughout this work, missing entries are marginalised in the previous way whenever present.

### 3.3.2 The generative model

In the Pleiades DANCe DR2 data set, cluster stars are mixed with field sources. Thus, cluster and field populations must be disentangled to achieve

the objective of this work. The probabilistic disentanglement of them, requires probabilistic models for each population, i.e. their likelihoods. The true values of these two likelihoods, field and cluster, are given by parametric relations embodying the state of knowledge of NYC. Using these likelihoods, individual cluster membership probability PDFs are computed (by means of Eq. 3.19) for each object in the data set. These PDFs together with a probability classification threshold, allow individual objects to be separated into cluster and field populations.

The inference process demands a set of  $N$  binary integers  $\mathbf{q}$ , one  $q_n$  for each object. The two possible values of these binary integers represent one of the two mutually exclusive possibilities: the object belongs to the cluster ( $q_n = 1$ ) or to the field population ( $q_n = 0$ ).

Let  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$  be the parameters of the cluster and field models, and  $p_c$  and  $p_f$  their likelihoods, respectively. Then, the likelihood of the data,  $\mathbf{D}$ , is,

$$p(\mathbf{D}|\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{n=1}^N p_c(\mathbf{d}_n|\boldsymbol{\theta})^{q_n} \cdot p_f(\mathbf{d}_n|\boldsymbol{\phi})^{(1-q_n)}. \quad (3.21)$$

The inference of these  $N$  binary integers will demand a computing power that is outside the current possibilities. Thus, instead of inferring them, I marginalise them using a prior probability, which is set in terms of a new and unique parameter  $\pi$ . It represents the *prior* probability that an object belongs to the field. Thus, the prior probability of  $\mathbf{q}$  is

$$p(\mathbf{q}|\pi) = \prod_{n=1}^N (1 - \pi)^{q_n} \cdot \pi^{(1-q_n)}. \quad (3.22)$$

Since  $\mathbf{q}$  represents the  $2^N$  possible ways to combine the binary variables  $q_n$ , then the marginalisation over all these possible states is,

$$\begin{aligned} p(\mathbf{D}|\pi, \theta, \phi) &= \sum_{\mathbf{q}} p(\mathbf{D}, \mathbf{q}|\pi, \theta, \phi) \cdot \Delta \mathbf{q} \\ &= \sum_{\mathbf{q}} p(\mathbf{D}|\mathbf{q}, \pi, \theta, \phi) \cdot p(\mathbf{q}|\pi) \cdot \Delta \mathbf{q} \\ &= \sum_{\mathbf{q}} \prod_{n=1}^N p_c(\mathbf{d}_n|\theta)^{q_n} \cdot p_f(\mathbf{d}_n|\phi)^{(1-q_n)} \cdot \prod_{n=1}^N (1 - \pi)^{q_n} \cdot \pi^{(1-q_n)} \cdot \Delta \mathbf{q} \\ &= \sum_{\mathbf{q}} \prod_{n=1}^N [(1 - \pi) \cdot p_c(\mathbf{d}_n|\theta)]^{q_n} \cdot [\pi \cdot p_f(\mathbf{d}_n|\phi)]^{(1-q_n)} \cdot \Delta \mathbf{q} \\ &= \prod_{n=1}^N (1 - \pi) \cdot p_c(\mathbf{d}_n|\theta) + \pi \cdot p_f(\mathbf{d}_n|\phi). \end{aligned} \quad (3.23)$$

This last equality is a rather complicated derivation which can be found in Press (1997),

and in Hogg et al. (2010a) for  $p_c$  and  $p_f$  in the exponential family. Also, a general derivation of this expression is given by Jaynes (2003). He obtains it assuming individual unknown probabilities  $p_n$  instead of  $q_n$  and marginalising over them with the aid of a prior.

Thus, the *generative model* or likelihood of the datum  $\mathbf{d}_n$  is

$$p(\mathbf{d}_n|\pi, \boldsymbol{\theta}_c, \boldsymbol{\theta}_f, \mathbf{u}_n) = \pi \cdot p_f(\mathbf{d}_n|\boldsymbol{\theta}_f, \mathbf{u}_n) + (1 - \pi) \cdot p_c(\mathbf{d}_n|\boldsymbol{\theta}_c, \mathbf{u}_n), \quad (3.24)$$

where  $\boldsymbol{\theta}_f$  and  $\boldsymbol{\theta}_c$  indicate the cluster and field parameters, while  $\mathbf{u}_n$  refers to the datum uncertainty. The probabilities  $p_f(\mathbf{d}_n|\boldsymbol{\theta}_f, \mathbf{u}_n)$  and  $p_c(\mathbf{d}_n|\boldsymbol{\theta}_c, \mathbf{u}_n)$  are the field and cluster models, respectively. These models are explained in detail in the next two sections.

In the following, I assume that the observed quantities, even if they contain missing values, resulted from the convolution of likelihood, given the *true* quantities, with the uncertainty process (see Section 3.1.1). The latter could be different for each object, thus resulting in different individual uncertainties. It means that the DANCe DR2 data set is modelled allowing for its intrinsic heteroscedasticity. Although these uncertainties are not assumed to have the same values, I do assume that they have the same shape. Thus, I assume that uncertainties in our data set belong to the multivariate normal family. This assumption is standard practice and is also supported by the large and heterogeneous origins of the DANCe DR2 data set.

### 3.3.3 The field population

To model the field population, I assume that the joint seven-dimensional probability distribution of the field can be factorised into the probability distributions of proper motions and photometry. Thus, the field likelihood of the proper motions and the photometry are assumed to be independent. Also, I assume that both distributions are described by Gaussian Mixture Models (GMM). The flexibility of GMMs to fit a variety of distribution geometries makes them a suitable model to describe the field density of the heterogeneous DANCe DR2 data set.

**A GMM is a probability distribution resulting from the linear combination of  $M$  Gaussian distributions,**

$$p_{GMM}(x|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{m=1}^M \pi_m \cdot \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \quad (3.25)$$

where  $\pi_m$  is the fraction of the  $m$ th Gaussian,  $\boldsymbol{\mu}_m$  its mean and,  $\boldsymbol{\Sigma}_m$  its covariance matrix. The  $m$  fractions must add to one.

Notice that the number of Gaussians in the mixture is not formally speaking a parameter, but rather it implies a collection of parameters. The number of these parameters increases linearly with the number of Gaussians and quadratically with the dimension.

According to Bouy et al. (2015), the number of Pleiades candidate members in the DANCe DR2 data set is 2010 (going up to 2109 in the combined Tycho-DANCe DR2) from a total of 1,972,245 sources . It means that the number of field objects dominates (99.9%) the DR2 data set. Even in our restricted  $10^5$  objects data set, RDR2 (see Sect. 2.7.2), the field still dominates with a 0.98 fraction. Thus, it can be assumed that any classification of candidate members will have a negligible impact on this figure. Therefore, it seems reasonable to assume that the GMM describing the field population can be frozen (fixed) during the process of cluster parameters inference. I elaborate more on this assumption.

The objects that Bouy et al. (2015) classified as belonging to the field are those whose cluster membership probabilities are lower than 0.75. This probability threshold corresponds to the one found by Sarro et al. (2014) after analysing the performance of their methodology when applied to synthetic data sets. In Sarro et al. (2014) the authors report that, at a probability threshold  $p = 0.75$ , the contamination and true positive rates are  $\sim 8\%$  and  $\sim 96\%$  respectively. Assuming that these values are correct, the real number of field objects would change by approximately 4% of the cluster members (adding the 8% of contaminants and subtracting the 4% of missed members), thus  $\sim 80$  objects. This value is negligible compared to the size of the data set ( $10^5$  objects in the RDR2). It represents the negligible fraction of  $8 \times 10^{-4}$ .

It can be further assumed that these hypothetically misclassified objects are spread in the space of observables. Indeed, these misclassified objects can be thought to have membership probabilities near the classification threshold. Thus, they may lay in the entanglement regions which corresponds, in proper motions, to a halo around the cluster centre, and in photometry, to a region around the cluster sequence. For example, these objects could have photometry that agrees with the cluster sequence but proper motions that are far from it. There are several combinations of observables values which may render these intermediate membership probabilities. The observable regions that these objects may populate allows us to assume that they are spread over the observable space.

If the misclassified objects are a few and spread over the observable space, then their contribution to the parameters of the GMM describing the field population can be neglected. Thus, the parameters of the GMM can remain fixed and out of the inference process.

The previous assumption is of paramount importance due to practical reasons. If I were to simultaneously infer the parameters of both cluster and field models, **even only those of the field proper motions GMM**, the required computing time would be excessive. The number of parameters in the field GMM goes up to  $\sim 300$ , more than twice that of the cluster model (85 parameters). **Even inferring only the 42 parameters of the field proper motions GMM would represent, at least, a 50% increase in**

**the computing time.** Instead, I first fit the parameters of the field GMM using the field objects in the RDR2 (the  $\sim 98,000$  objects with membership probabilities below 0.75) and then I keep these parameters fixed in the inference process.

The number of gaussians for the proper motions and photometric GMM is found using the Bayesian Information Criterion (BIC, [Schwarz 1978](#)). The BIC is a model selection criterium that aims at avoiding over-fitting. It represents a compromise between the likelihood,  $\mathcal{L}$ , of the  $n$  data points, and the number of parameters,  $k$ . This is,

$$BIC = \ln n \cdot k - 2 \ln \mathcal{L}. \quad (3.26)$$

To estimate the parameters of the GMM we use the Expectation Maximisation (EM) algorithm. However, the missing values in the photometry prevent the use of the standard form of the algorithm (see for example Chapter 9 of [Bishop 2006](#)). Instead, I estimated these parameters with the modified version of the EM algorithm for GMM found by [McMichael \(1996\)](#) and rediscovered by me. On this version, objects with missing values also contribute to estimate the maximum-likelihood (ML) parameters. I applied this algorithm to GMM whose number of components ranged from 1 to 20. The optimal number of gaussians suggested by the BIC is 14.

**In the case of proper motions, since they do not contain missing values, I computed the GMM parameters with the standard EM algorithm. The BIC finds a model with 15 components, the majority of them with large variances and small fractions (see Fig. 3.3a).** These small-fraction Gaussians fit an extended component of the proper motions distributions, which is clearly non-Gaussian. For this reason, we decided modify the mixture of Gaussians by adding a uniform distribution  $\mathcal{U}(\mathbf{d}_{pm}|S_{\mu_\alpha}, S_{\mu_\delta})$ , with  $S_{\mu_\alpha}$  and  $S_{\mu_\delta}$  the support of the proper motions in Right Ascension and Declination, respectively (see Table 2.2). I changed the EM algorithm to properly account for this modification. The BIC applied to this new mixture renders more reasonable results (see Fig. 3.3b), with an eight components mixture distribution: seven Gaussians plus the uniform. This modification improves the GMM while simultaneously reduces the number of parameters.

Finally, the field likelihood  $p_f(\mathbf{d}|\boldsymbol{\theta}_f, \mathbf{u})$  of an object with measurements  $\mathbf{d}$ , given its standard uncertainties  $\mathbf{u}$  and the field parameters,  $\boldsymbol{\theta}_f$ , is

$$p_f(\mathbf{d}|\boldsymbol{\theta}_f, \mathbf{u}) = \left[ \pi_{f,pm,0} \cdot \mathcal{U}(\mathbf{d}_{pm}|S_{\mu_\alpha}, S_{\mu_\delta}) + \sum_{i=1}^7 \pi_{f,pm,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{f,pm,i}, \boldsymbol{\Sigma}_{f,pm,i} + \mathbf{u}_{pm}) \right] \cdot \left[ \sum_{i=1}^{14} \pi_{f,ph,i} \cdot \mathcal{N}(\mathbf{d}_{ph}|\boldsymbol{\mu}_{f,ph,i}, \boldsymbol{\Sigma}_{f,ph,i} + \mathbf{u}_{ph}) \right]. \quad (3.27)$$

The first and second brackets represent the proper motion (subindex  $pm$ ) and photometric

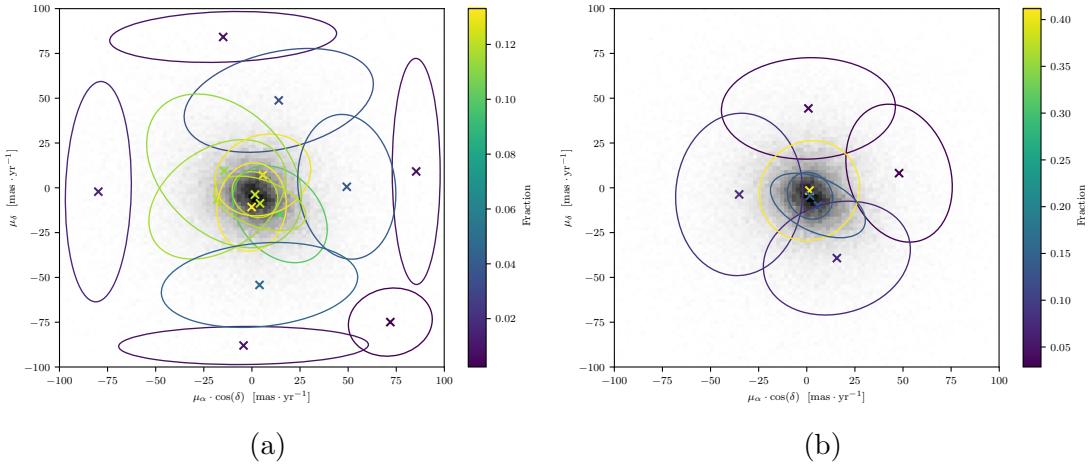


Figure 3.3: Gaussian (a) and Modified (b) Mixture Models fitted to the proper motions of the field objects in the RDR2 data set (density in grey scale). The crosses and ellipses indicate the means and the one- $\sigma$  covariance matrices, respectively. The colour code indicates the value of the fraction in the mixture. Uniform distribution not shown in (b)

(subindex  $ph$ ) models, respectively, with  $\pi_f, \mu_f, \Sigma_f$  their fractions, means and covariance matrices, respectively. **The first term of the proper motion model is the uniform distribution, with  $\pi_{f,pm,0}$  its fraction.** The uncertainty process is convolved (see Section 3.1) with the assumed models using the individual proper motions ( $\mathbf{u}_{pm}$ ) and photometric ( $\mathbf{u}_{ph}$ ) uncertainties.

### 3.3.4 The cluster population

Similarly to what I assume for the field population model, I also assume that the cluster model or likelihood can be factorised into the product of the proper motions distribution times the photometric distribution. Thus, I assume these two models are independent.

It is known that unresolved systems of stars (groups of stars that, given the spatial resolution of the telescope, are seen as an individual object) have an increased brightness proportional to the multiplicity of the system. In particular, if an unresolved system is made of two equally luminous objects, then its magnitude is 0.752 times brighter than that of an individual object. This happens for equal mass binaries.

Since the pioneer work of Trumpler (1921), we know that some of the Pleiades members are double systems. Recently, Sarro et al. (2014) show evidence that some of these double systems lie in an equal-mass binaries (EMB) sequence. Those authors model this EMB sequence assuming that its number of objects is 20% of the total number of cluster members. In the present work, we also model objects in this displaced EMB sequence but we do not assume that its proportion is fixed; we infer it from the data.

Unresolved multiple systems, binary systems particularly, have an impact on the cluster proper motions distribution. In stellar clusters, the massive objects are expected to fall towards the centre of the gravitational potential at a higher rate than that of

the less massive ones (see for example [Binney & Tremaine 2008](#), p. 556). This behaviour arises from stellar encounters in which the energy exchange results in the less massive objects gaining kinetic energy and the massive ones losing it. **Since binaries and multiple systems are typically more massive than the average object, they are expected to fall towards the cluster centre.** From an astronomical point of view, an unresolved system shifts the photo-centre of its images when compared to that of a single object. Given the previous considerations, we decide to model the EMB as an independent population in the proper motions. Furthermore, we pair this proper motions model to its photometric counterpart. This more comprehensive statistical model allows us to directly compare the kinematic and photometric properties of the EMB population with those of the rest of the cluster.

For the sake of simplicity, in the following whenever I refer to the photometric or proper motions model of the EMBs, I call it the EMB sequence model (with the subindex  $B_s$ ). I refer to the model of the rest of the stars as the cluster sequence model or simply single stars. Despite that this is an abuse of the terminology, because there are binaries and multiple systems with different mass ratios, it keeps the text more readable.

### Photometric model of single and EMB stars

The cluster photometric sequences, both for single and EMB stars, are modelled with cubic splines, one for each of the  $Y, J, H, K_s$  vs  $CI$  CMDs. I choose the spline series due to their fitting properties. I tried several polynomial bases (monic, Laguerre, Hermite, Chebyshev) but in spite of the order, they were not able to fit the sequence. In specific, they do not fit the region around  $CI \sim 3$  where the slope is higher.

Despite their superior flexibility, spline series require more parameters than polynomials. In addition to the coefficients of the series, they require the specification of points called knots. These knots represent the starting and ending points of the spline sections.

Any spline function can be represented in terms of basis-splines or B-splines. By definition, a B-spline of order  $n$  is a piece-wise polynomial function of order  $n$  in the interval  $t_0 \leq x \leq t_n$ . The boundary and internal points,  $\mathbf{t} = \{t_0, t_1, \dots, t_n\}$  represent the knots. For a given set of knots, there is one and only one B-spline representation of the spline, thus the name basis spline. In particular, any cubic spline can be represented as,

$$S_3(CI, \boldsymbol{\beta}, \mathbf{t}) = \sum_i \beta_i \cdot B_{i,3}(CI, \mathbf{t}). \quad (3.28)$$

Where  $B_{i,n}$  are the B-splines given by the Cox-de Boor recursive formula, and  $\boldsymbol{\beta}$  are the coefficients of the series. For more details on splines and the Cox-de Boor formula see [De Boor \(1978\)](#).

Despite their fitting properties, B-splines **present a problem** when inferring simultaneously their coefficients and knots: there is multi-modality in the parametric space

(Lindstrom 1999). It means that at least more than one combination of parameters produces the same solution. To avoid this multi-modality, I decided to keep the knots fixed throughout the inference. Although this decision reduces the flexibility of the splines, it allows a still better fit than that of the tested polynomials. To obtain the maximum-likelihood (ML) estimate of the knots I use the algorithm of Spiriti et al. (2013). This algorithm, implemented in the *freetknotsplines* R package, allows to simultaneously obtain the knots and the best truncation value for the spline series. It uses the BIC to select among competing models. In order to obtain both the truncation of the series and the value of the knots, I use the candidate members of Bouy et al. (2015). The BIC indicates that seven coefficients is the best number of components for the B-spline series, with the knots at  $\mathbf{t} = \{0.8, 3.22, 3.22, 5.17, 8.0\}$ . I tested different number of knots, ranging from two to nine, with five the best configuration given by the BIC. **In general, the continuity of a spline function is  $C^{p-k}$ , with  $p$  the degree of the spline, and  $k$  the highest multiplicity of the knots (De Boor 1978).** In our case, since the knot at 3.22 has a multiplicity of two, then the resulting spline has lost one degree of continuity. It is now  $C^1$  continuous. It means that the spline and its derivative are continuous, thus ensuring a smooth function.

As I mentioned in the introduction to this Section, we assume that the observed photometric quantities are drawn from a distribution resulting from the convolution of the observed uncertainties, with the likelihood, given the *true* quantities. We also assume that the model has an intrinsic dispersion that addresses photometric variations resulting from astrophysical phenomena not treated by our model (see Section 3.1.1). These phenomena include, but are not limited to, age, metallicity and distance dispersions, unresolved systems (other than EMB), variability, and transits. If we were to assume no *true* intrinsic dispersion, then any deviation from the *true* quantities would have to be explained *only* on the basis of the observational uncertainties. Thus it would result in an over-simplistic model, which would underestimate the likelihood of hypothetical true cluster members.

This photometric dispersion shows an skewed distribution (see Figure 2 of Hillenbrand et al. 2008, which I reproduce in Fig. 3.4), which we model with a multivariate skewed normal distribution (CSN, see for example González-Farias et al. 2004; Gupta et al. 2004). Despite the fact that the CSN requires only five parameters more than a multivariate normal distribution, the computing of its density takes  $\sim 50\%$  more CPU time than that of the multivariate normal distribution (MND). Due to computational constraints, we decided to postpone the results of the CSN until our computational resources allow it.

Instead of using the CSN, we model the intrinsic photometric dispersion of both the cluster and EMB sequences with a MND. It has five dimensions corresponding to our photometric reference set:  $CI, Y, J, H, K_s$ . The B-splines model the *true* photometric quantities, both for the cluster sequence,  $\mathbf{t}_{ph;Cs}$ , and the EMB,  $\mathbf{t}_{ph;Bs}$ . The latter is displaced 0.75 into the bright side of the cluster sequence. In the following, the matrix

$\Sigma_{clus}$ , represents the covariance matrix of this MND. By definition, covariance matrices are symmetric and positive semi-definite. Therefore, from the 25 entries in  $\Sigma_{clus}$ , only 15 are unique. These are also inferred from the data set.

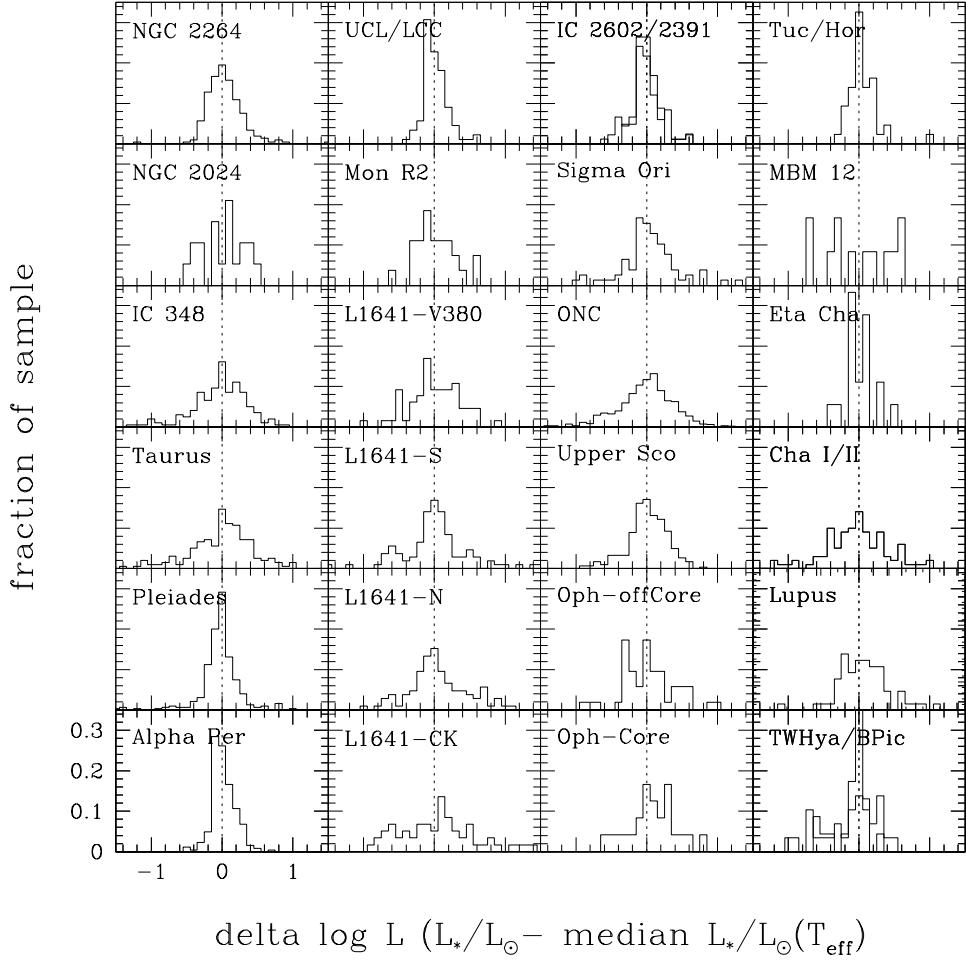


Figure 3.4: Histogram of luminosity dispersion for young clusters. Reproduced from Figure 2 of Hillenbrand et al. (2008), *An Assessment of HR Diagram Constraints on Ages and Age Spreads in Star-Forming Regions and Young Clusters*, Astronomical Society of the Pacific Conference Series, Vol. 384.

Thus, the *true* photometry is given by,

$$\begin{aligned}\mathbf{t}_{ph;Cs} &= \{CI, Y, J, H, K_s\}, \\ \mathbf{t}_{ph;Bs} &= \{CI, Y - 0.75, J - 0.75, H - 0.75, K_s - 0.75\},\end{aligned}$$

where

$$\begin{aligned} Y &= \mathcal{S}_Y(CI, \hat{\beta}_Y), \\ J &= \mathcal{S}_J(CI, \hat{\beta}_J), \\ H &= \mathcal{S}_H(CI, \hat{\beta}_H), \\ K_s &= \mathcal{S}_{K_s}(CI, \hat{\beta}_{K_s}), \end{aligned}$$

with  $\hat{\beta}_i, i \in \{Y, J, H, K_s\}$  the vectors of seven coefficients of the B-splines for the  $Y, J, H, K_s$  bands. For the sake of simplicity I denote this  $4 \times 7$  coefficients matrix as  $\boldsymbol{\beta}$ .

Since the photometry of the EMB is a linear transformation,  $T_{Bs}$ , of the mean *true* photometry of cluster sequence, no extra parameters are required. Therefore,

$$\mathbf{t}_{ph;Cs} = \mathcal{S}(CI, \boldsymbol{\beta}) \quad (3.29)$$

$$\mathbf{t}_{ph;Bs} = T_{Bs}(\mathcal{S}(CI, \boldsymbol{\beta})). \quad (3.30)$$

Thus, cluster and EMB likelihoods of an object with photometric measurements  $\mathbf{d}_{ph}$ , and standard uncertainties  $\mathbf{u}_{ph}$ , are:

$$\begin{aligned} p_{Cs}(\mathbf{d}_{ph}|CI, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) &= \mathcal{N}(\mathbf{d}_{ph}|\mathbf{t}_{ph;Cs}, \mathbf{u}_{ph} + \Sigma_{clus}), \\ p_{Bs}(\mathbf{d}_{ph}|CI, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) &= \mathcal{N}(\mathbf{d}_{ph}|\mathbf{t}_{ph;Bs}, \mathbf{u}_{ph} + \Sigma_{clus}), \end{aligned} \quad (3.31)$$

where  $\mathbf{t}_{ph;Cs}$  and  $\mathbf{t}_{ph;Bs}$  are given by Equations 3.29 and 3.30, respectively.

Since the splines are parametrised by the true  $CI$  of each object, we have more parameters than objects in our data set <sup>11</sup>. This *true*  $CI$  is unknown even if its observed value is not missing. We solve this problem (it is a computational problem!) by marginalising these nuisance parameters.

To marginalise these  $CIs$  we need a prior, which we provide in a hierarchical way (thus the name Bayesian Hierarchical model). This marginalisation leaves behind a precise estimate of the parameters of the prior distribution. Paradoxically, all objects, even those without a measurement of the  $CI$ , contribute to this estimate. Here lies the force of the BHM.

We model the prior of the *true*  $CI$  as a truncated ( $0.8 \leq CI \leq 8$ ) univariate GMM with five components, whose parameters are also inferred from the data. We choose five components as suggested by the BIC computed from the EM algorithm for GMM applied to the candidate members of Bouy et al. (2015). I tested larger number of components in the mixture (up to ten), but the posterior distribution did not changed significantly, thus indicating that the

---

<sup>11</sup>Although this sounds crazy, the rules of probability calculus do not discard this possibility.

### BIC value was a proper assumption.

The GMM modelling the prior of the *true CI* is

$$p(CI|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}) = \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}). \quad (3.32)$$

In this last Equation, the symbol  $\mathcal{N}_t$  stands for the truncated ( $0.8 < CI < 8$ ) univariate normal distribution.

Then, the marginalisation of  $CI$  runs as follows:

$$\begin{aligned} p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) &= \int p_{Cs}(\mathbf{d}_{ph}, CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \\ &= \int p_{Cs}(\mathbf{d}_{ph}|CI, \boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot p_{Cs}(CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \end{aligned} \quad (3.33)$$

$$\begin{aligned} p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) &= \int p_{Bs}(\mathbf{d}_{ph}, CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \\ &= \int p_{Bs}(\mathbf{d}_{ph}|CI, \boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot p_{Bs}(CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI. \end{aligned} \quad (3.34)$$

In these Equations,  $\boldsymbol{\theta}_c$  stands for all cluster parameters related to photometry, and the first and second terms of the integrals in the last equalities correspond to Equations 3.31 and 3.32, respectively. The distribution of  $CI$  depends only on  $\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}$ , thus, the cluster and equal-mass binaries likelihoods of datum  $\mathbf{d}_{ph}$  are

$$\begin{aligned} p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\ &= \int \mathcal{N}(\mathbf{d}_{ph}|\mathcal{S}(CI, \boldsymbol{\beta}), \mathbf{u}_{ph} + \Sigma_{clus}) \cdot \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \cdot dCI \\ p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\ &= \int \mathcal{N}(\mathbf{d}_{ph}|T_{Bs}(\mathcal{S}(CI, \boldsymbol{\beta})), \mathbf{u}_{ph} + \Sigma_{clus}) \cdot \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \cdot dCI. \end{aligned} \quad (3.35)$$

The observed  $CI$  and magnitudes help us to reduce the computing time of the marginalisation integral. We use them to discard regions of the integral in which the argument is almost zero (i.e. far from the measured values). Although we allow the nuisance parameters  $CIs$  to have all their possible values, the data, by means of the likelihood, give us information about the distribution of these individual nuisance parameters. To use this information, we proceed as follows. First, we compare the observed photometry to the true one (i.e. the cluster sequence given by the splines). For it we use a grid of 300 points uniformly distributed in the domain of  $CI$  ( $0.8 < CI < 8$ )<sup>12</sup>. Then, we find the

---

<sup>12</sup>As explained in Section 2.7.2, if this  $CI$  range will have covered that of all the objects in the DANCe DR2 data set, the marginalisation integral would have to be computed over the extended  $CI$  range.

point,  $p$ , of the grid that is closest to the vector of the observed photometry. Distance is computed under the Mahalanobis metric. This metric takes into account the observational uncertainty,  $\mathbf{u}_{ph}$ , and the intrinsic dispersion of the cluster sequence,  $\Sigma_{clus}$ . Finally, the limits of the marginalisation integral are defined as those given by a ball of 3.5 Mahalanobis distances around point  $p$ . Contributions outside this ball are negligible to the integral ( $< 4 \times 10^{-4}$ ).

### Proper motion model of EMB and single stars

As mentioned before, we assume that the cluster population has two subpopulations: single and EMB stars. We model the proper motions of these two subpopulations with independent GMM. If the cluster is virialised (see Chapter 2), we can assume that **the distribution of its velocity modulus** is almost Maxwellian (Maxwell-Boltzman distribution). Therefore a GMM is a reasonable **approximation**. Furthermore, in the absence of external forces, a virialised system is expected to have spherical symmetry both in its spatial and velocity distributions. Thus we can safely assume that the gaussians within each GMM are concentric, thus they share the same mean. However we allow independent means for both single and EMB subpopulations. The assumption of spherical symmetry may be a weak one in the presence of the galactic potential. It can perturb the cluster and deviate its spatial and velocity distribution from spherical symmetry. Furthermore, the ellipticity of the spatial distribution, which has been reported to be no-negligible ( $\epsilon = 0.17$ , according to [Raboud & Mermilliod 1998](#)), can be due to projection effects that further deviate the observed velocity distribution profile from spherical symmetry. Nevertheless, since we model the covariance matrices of the GMM of both single and EMB, as full covariance matrices, any departure from the spherical symmetry in the velocity distribution can still be modelled by the non-diagonal entries of these matrices.

We infer the parameters of these GMMs as part of our Bayesian hierarchical model. However we set a priori the number of gaussians in each GMM. Not doing so will demand a technique in which the model parameters can be augmented. Although such techniques already exist, they are still under computational development (see [Fan & Sisson 2011](#), for a review of reversible jump MCMC).

Using the EM algorithm for GMM and the proper motions of the candidate members of [Bouy et al. \(2015\)](#), I obtained the ML estimates for the GMM likelihoods. I did this for configurations of GMM ranging from one to five components. The BIC (Eq. 3.26) suggested four and two components for the cluster and EMB GMMs, respectively. Since covariance matrices are always symmetric, only three parameters are needed to fully specify the covariance matrices of these bivariate normal distributions.

The cluster (subindex  $Cs$ ) and EMB (subindex  $Bs$ ) likelihoods of an object with proper

---

Therefore, the computing time of the BHM would have also increased proportionally to the number of extra points in this integral.

motions measurements  $\mathbf{d}_{pm}$ , and uncertainties  $\mathbf{u}_{pm}$ , are

$$\begin{aligned} p_{Cs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Cs}, \boldsymbol{\mu}_{Cs}, \boldsymbol{\Sigma}_{Cs}, \mathbf{u}_{pm}) &= \sum_{i=1}^4 \pi_{Cs,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{Cs}, \Sigma_{Cs,i} + \mathbf{u}_{pm}) \\ p_{Bs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Bs}, \boldsymbol{\mu}_{Bs}, \boldsymbol{\Sigma}_{Bs}, \mathbf{u}_{pm}) &= \sum_{i=1}^2 \pi_{Bs,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{Bs}, \Sigma_{Bs,i} + \mathbf{u}_{pm}). \end{aligned} \quad (3.36)$$

Finally, combining the proper motions and photometric models, the total cluster likelihood of an object with measurement  $\mathbf{d}$ , and uncertainties  $\mathbf{u}$ , is

$$\begin{aligned} p_c(\mathbf{d}|\boldsymbol{\theta}_c, \mathbf{u}) &= \pi_{CB} \cdot p_{Cs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Cs}, \boldsymbol{\mu}_{Cs}, \boldsymbol{\Sigma}_{Cs}, \mathbf{u}_{pm}) \cdot p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\ &\quad + (1 - \pi_{CB}) \cdot p_{Bs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Bs}, \boldsymbol{\mu}_{Bs}, \boldsymbol{\Sigma}_{Bs}, \mathbf{u}_{pm}) \cdot p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}), \end{aligned} \quad (3.37)$$

where  $\pi_{CB}$  is the parameter representing the proportion or fraction of single cluster sequence stars in the single-EMB mixture model. The photometric and proper motions likelihoods are given by Equations 3.35 and 3.36, respectively.

Finally, before ending this Section, I present a summary of the symbols used to represent the parameters in the BHM (Table 3.1), together with its graphical representation in the form of a PGM (Fig. 3.5).

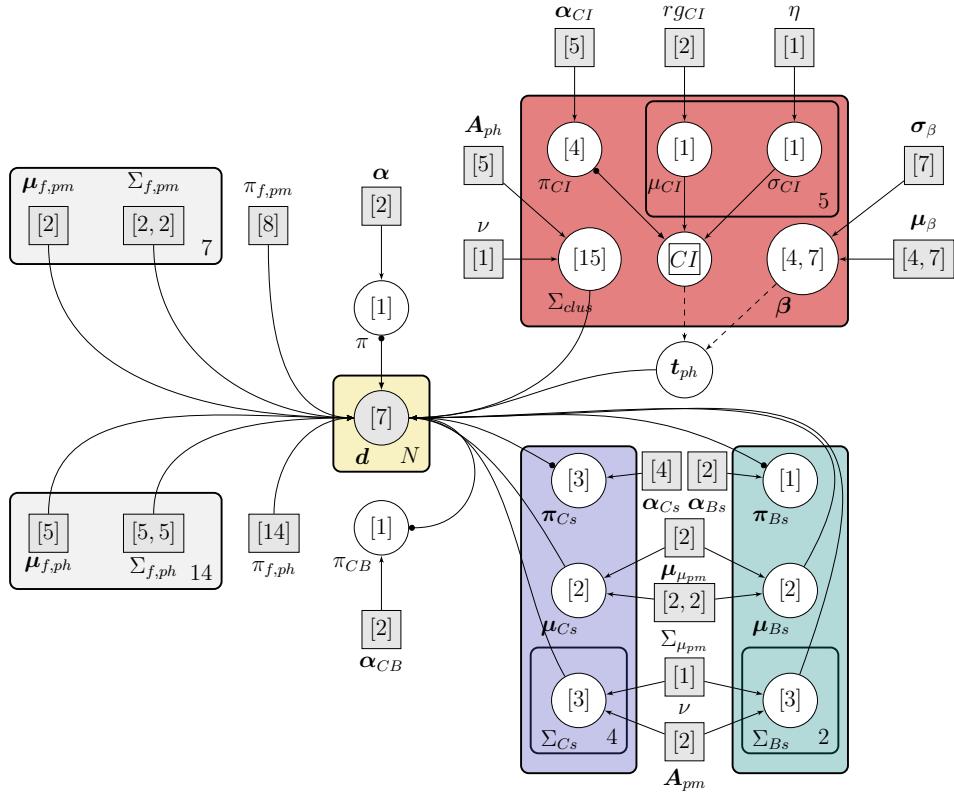


Figure 3.5: Probabilistic graphical model representing the BHM. The left grey plates show the field model. The middle yellow plate shows the node where the likelihood is computed for each datum,  $\mathbf{d}$ . The right plates describe the relations among parameters in the cluster model. The photometric cluster model (red) is on top, while the proper motions cluster (blue) and equal-mass binaries (green) are at the bottom left and right, respectively. See Section 3.2.3 for more details. Reproduced from Figure 19 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, Vol. submitted.

### 3.4 The methodology for the analysis of the PSD

In this Section, I present the methodology applied to investigate the Projected Spatial Distribution (PSD). Although this analysis uses the cluster membership probabilities recovered by the BHM, is independent of the latter. It means that the kinematic and photometric distributions derived by the BHM are assumed to be independent of the PSD of the cluster. Section 2.7.1 details the reasons of this decision.

The generative model of the PSD is similar to that of the BHM ( Eq. 3.24). However, instead of inferring the fraction of field objects in the data set (which would be similar to the estimated contamination rate on the HMPS, see Section 4.1), each object contributes to the cluster model proportionally to its membership probability,  $P$  (the ones from the BHM). Thus the generative model of the PSD for object  $n$  is

$$p(\mathbf{d}_n, P_n | \boldsymbol{\theta}_c, \boldsymbol{\theta}_f) = (1 - P_n) \cdot p_f(\mathbf{d}_n | \boldsymbol{\theta}_f) + P_n \cdot p_c(\mathbf{d}_n | \boldsymbol{\theta}_c), \quad (3.38)$$

with  $\mathbf{d}_n$  its vector of observations, comprising the right ascension  $\alpha$  (R.A.) and the declination  $\delta$  (Dec.), and  $P_n$  its cluster membership probability. The terms  $p_f(\mathbf{d}_n|\boldsymbol{\theta}_f)$  and  $p_c(\mathbf{d}_n|\boldsymbol{\theta}_c)$  are the field and cluster likelihood, respectively, of the datum  $\mathbf{d}_n$  given the field,  $\boldsymbol{\theta}_f$  and cluster,  $\boldsymbol{\theta}_c$ , parameters. Since all objects in the Pleiades DANCe DR2 have R.A. and Dec. uncertainties which are symmetric, almost identical, and negligible ( $\sim 2 \times 10^{-6}$  deg) compared to the size of the cluster ( $\sim 7^\circ$ ), we can safely assume that: i) the data are independent and identically distributed, and ii) the observed positions correspond to the *true* ones. Thus, the uncertainties are not included in the generative model.

We would expect that the fraction of contaminants in the HPMS will be uniformly distributed in right ascension and declination because the position on the sky was explicitly removed from the calculation of membership probabilities. Except perhaps for a mild positive gradient of the density towards the Galactic centre. Hence, we model these contaminants with a uniform spatial distribution  $\mathcal{U}$ . In this way, the contribution of these hypothetical contaminants will not bias our inference of the cluster parameters.

In the following Section, I describe the kinds of PSD models that we included in our model selection analysis: those with radial symmetry, and those with luminosity segregation (a proxy for mass segregation). The former are the classical ones in the literature. The latter allow us to put a solid statistical background to the assumption of mass segregation in the Pleiades cluster (see for example [Moraux et al. 2004](#); [Converse & Stahler 2010](#)) without relying on mass-luminosity transformations for individual objects.

### 3.4.1 Models with radial symmetry

The first alternative is King profile ([King 1962](#)), which has been widely used in the Pleiades (see Section 2.3) and to describe open clusters (see [Alonso-Santiago et al. 2017](#); [Panwar et al. 2017](#), for recent applications), globular clusters ([Myeong et al. 2017](#)) and even to study galaxies ([Robotham et al. 2017](#)), halo substructure ([Sohn et al. 2007](#)) and the dark matter distribution ([Jiang & van den Bosch 2016](#)).

The analytical description of the surface number density of stars counts  $n$  is given by

$$\rho(R) = k \cdot \left( \frac{1}{\sqrt{[1 + (R/R_c)^2]}} - \frac{1}{\sqrt{[1 + (R_t/R_c)^2]}} \right)^2, \quad (3.39)$$

where  $R_c$  (the core radius) is a scale factor,  $R_t$  is the tidal radius, and  $k$  is a constant related (but not equal) to the central surface density. In the following we use  $R$  instead of  $r$  (as is often common in some of the articles that introduced the profile analytic expressions used here) to refer to the distance from the system centre projected on the celestial sphere. Also, we use  $\rho$  to denote the surface number density of stars, and,  $I$ , to refer to the projected surface brightness.

We have also considered the model proposed by [Elson et al. \(1987\)](#) (henceforth EFF) to describe young open clusters in the Large Magellanic Cloud. Their surface density (in

star counts per solid angle) is given by

$$\rho(R) = \rho(0) \cdot (1 + (R^2/A^2))^{\frac{\gamma}{2}}. \quad (3.40)$$

Equation 3.40 is introduced heuristically and hence, the two parameters  $A$  and  $\gamma$  have no physical meaning (although  $A$  must have the same units as  $R$ ).

Finally, we analyse a more general parameterisation introduced in Lauer et al. (1995), Byun et al. (1996) and Zhao (1997), where the projected mass density is given as

$$\rho(R) = \frac{k'}{(R/R_b)^\gamma \cdot (1 + (R/R_b)^{1/\alpha})^{(\gamma-\beta)\alpha}}. \quad (3.41)$$

Equation 3.41 represents a double power law with  $\gamma$  being the exponent in the innermost region,  $\beta$  the exponent in the outer region,  $\alpha$  is the width of the transition region,  $R_b$  is the so called break radius, and  $k'$  is a scale constant. In purity, the aforementioned works assume this functional form for the projected surface brightness  $I$ , the projected mass density  $\rho$ , and for the volume density  $v$ . Although the latter two are related by integration:

$$\rho(R) = \int_0^\infty v(r) \cdot dz \quad (3.42)$$

where  $z$  is the distance along the line of sight corresponding to a radial distance projected onto the plane of the sky  $R$ . Volume In this work we will assume that the analytical expression in Eq. 3.41 is correctly stated in terms of the number density and not of volume density.

This model is a more general analytical expression that comprises many simpler models each of which represent particular choices of the model parameters. For example, the model put forward by Plummer (1911) to describe the projected spatial density corresponds to the general model with  $\alpha = 1/2$ ,  $\beta = 5$  and  $\gamma = 0$ . Several density profiles proposed to describe galaxies can be grouped by particular choices of the parameters. For example,  $\alpha = 1$  includes models by Navarro et al. (1997), Hernquist (1990), Jaffe (1983), and Moore et al. (1999), and  $\alpha = 1/2, \gamma = 0$  includes the aforementioned model by Plummer (1911), and also models by Sackett & Sparke (1990) and de Zeeuw (1985). King's profile, however, cannot be cast into this general model unless the tidal radius  $R_t$  is fixed at infinity. This is an important difference that will nevertheless have little effect on our comparisons for reasons that will become apparent further below.

In all three cases, we need to convert the projected stellar densities (number of stars per unit solid angle) into probability density functions that describe the probability of finding a star between  $R$  and  $R + dR$  under the assumption of spherical symmetry. In all three cases we can construct the probability densities from their definitions:

$$\rho(R) = \frac{N \cdot p(R) \cdot dR}{2\pi \cdot R \cdot dR} \quad (3.43)$$

where  $N$  is the total number of stars in the system. Applying Equation 3.43 to King's profile, we obtain

$$p(R) = \frac{k \cdot 2\pi}{N} \cdot R \cdot \left( \frac{1}{\sqrt{1 + (R/r_c)^2}} - \frac{1}{\sqrt{1 + (r_t/r_c)^2}} \right)^2. \quad (3.44)$$

Actually, in probabilistic inference we write this probability function as:

$$p(R|r_c, r_t, k_1, \mathcal{M}_1) = k_1 \cdot R \cdot \left( \frac{1}{\sqrt{1 + (R/r_c)^2}} - \frac{1}{\sqrt{1 + (r_t/r_c)^2}} \right)^2, \quad (3.45)$$

where we have defined a new constant  $k_1 = \frac{k \cdot 2\pi}{N}$ , and made explicit the dependence of the probability on the underlying analytical expression ( $\mathcal{M}_1$ ), and the values of the parameter set ( $k_1, r_c$ , and  $r_t$ ). In practice  $k_1$  is treated as a normalisation constant (to enforce unit integral) and there is no need to know the total number of stars in the system. This PDF corresponds to the cluster model of *likelihood*.

Likewise, the expression for the EFF model is

$$p(R|a, \gamma, k_2, \mathcal{M}_2) = k_2 \cdot R \cdot (1 + (R^2/a^2))^{\frac{\gamma}{2}}. \quad (3.46)$$

And finally, the generic model will be given by

$$p(R|\alpha, \beta, \gamma, R_b, k_3, \mathcal{M}_3) = \frac{k_3 \cdot R}{(R/R_b)^\gamma \cdot (1 + (R/R_b)^{1/\alpha})^{(\gamma-\beta)\alpha}}. \quad (3.47)$$

In all three cases, the  $R$  coordinate is defined with respect to the origin of a coordinate axis. The actual values of  $R$  then depend on this choice of origin. We therefore infer this centre as part of our models in order to incorporate all sources of uncertainty.

### Extensions of the classical profiles

In addition to the classical profiles we have tested two extensions of the King profile, and one variant of the generalised profile. Let us describe them in order.

We define the generalised King profile (hereafter GKing profile) as the classical King profile without fixing the exponents of the analytical expression. Instead of Equation 3.45, we have

$$p(R|r_c, r_t, k_1, \mathcal{M}_1) = k_1 \cdot R \cdot \left( \left( 1 + (R/r_c)^{\frac{1}{\alpha_1}} \right)^{-\alpha_1} - \left( 1 + (r_t/r_c)^{\frac{1}{\alpha_1}} \right)^{-\alpha_1} \right)^{\alpha_2} \quad (3.48)$$

where the classical King's profile is recovered for  $\alpha_1 = 0.5$  and  $\alpha_2 = 2$ . Robotham et al. (2017) use a modified empirical King profile that is similar to the one presented here. However, their profile is more restrictive and requires that  $\alpha_2 = \alpha_1^{-1}$  and both terms  $(R/r_c)$ , and  $(r_t/r_c)$  will be at the power of 2.

The optimised/generalised King profile (hereafter OGKing) is the GKing profile with the values of  $\alpha_1$  and  $\alpha_2$  fixed so as to maximise the evidence evaluated with all DANCe DR2 sources. The reasons for this will become apparent after Section 4.5, but we can write here that this choice is the most favourable for the GKing profile in the context of model selection. These values are  $\alpha_1 = 0.4$  and  $\alpha_2 = 1.2$ . Notice that  $\alpha_1$  is very similar to the King  $\alpha_1 = 0.5$ , which indicates that the central region is almost identical to King's profile. On the other hand,  $\alpha_2 = 1.2$  is lower than King's  $\alpha_2 = 2$ , thus showing that the main difference comes from the outskirts of the cluster. More details will be given in Section 4.5.

Finally, the restricted/generalised profile (RGP) corresponds to the generalised profile with the value  $\gamma$  fixed at 0.

### 3.4.2 Central symmetry constraint

So far, we have defined a set of three models (King, EFF and the generalised profile GP) and their extensions. Each has a different set of parameters. King's model depends on two parameters ( $r_c$  and  $r_t$ ); EFF's model depends on two parameters ( $a$  and  $\gamma$ ); the generalised profile depends on four parameters ( $\alpha$ ,  $\beta$ ,  $\gamma$  and  $R_b$ ).

In reality, there are always two more parameters that do not appear explicitly in any of the classical analytical formulations of the King, EFF or generalised profiles. These are the cluster centre coordinates from which all radial distances  $R_i$  are measured. It is not a minor question because the problem is degenerate, and there is a maximum likelihood solution for each choice of the cluster centre. In principle, one could even choose a poor cluster centre estimate that renders the angular distribution of members asymmetric, and obtain a maximum likelihood fit better than those obtained with a better centred estimate. The models assume central symmetry, but this can only be ensured approximately. There is a region of non-negligible extent, where the cluster centre may be, and any particular choice of its position will influence the posterior distribution inferred. Thus, in order to propagate appropriately this uncertainty about the cluster centre position in our posterior inferences, we have included the two cluster centre coordinates,  $\alpha_c$  and  $\delta_c$ , and the polar angle  $\eta$  as further parameters of our models. The polar angle  $\eta$  ensures that the resulting symmetry does not depend on the RA and Dec. coordinate system. Furthermore, this angle will be used in the elliptical models of the next section, where the position angle is computed from the semi-major axis. Thus, the use of the same likelihood functions for the centre coordinates will allow us to perform a fair comparison of the radial models, which is the main objective of this section.

For any given choice of the cluster centre coordinates  $\alpha_c$ ,  $\delta_c$  and polar angle  $\eta$ , we calculate the radial distance,  $R(\alpha_c, \delta_c)$ , and the position angle  $\phi(\eta)$  of each star in our data set. To avoid biases introduced by projection effects of objects located far from the cluster centre, we project each object coordinates into the plane of the sky along the line-of-sight

vector (see for example, Eq. 1 of [van de Ven et al. 2006](#)). These projected coordinates are

$$\begin{aligned}\tilde{x} &= \sin(\alpha - \alpha_c) \cdot \cos(\delta) \\ \tilde{y} &= \cos(\delta_c) \cdot \sin(\delta) - \sin(\delta_c) \cdot \cos(\delta) \cdot \cos(\alpha - \alpha_c)\end{aligned}\quad (3.49)$$

From these projected coordinates, the radial distance  $R$  and the position angle  $\phi$  are computed as

$$\begin{aligned}R &= \sqrt{\tilde{x}^2 + \tilde{y}^2}, \\ \phi &= \arctan 2(\tilde{x}, \tilde{y})\end{aligned}\quad (3.50)$$

The requirement of central symmetry is enforced by the inclusion of a multiplicative term in the likelihood. For a given set of parameters values of  $\alpha_c, \delta_c, \eta$  we divide the computed polar angles of individual stars into four equally spaced quadrants (with separations at  $\phi = \{0, 90, 180, 270\}^\circ$ ) and require that the number of stars in each quadrant be Poisson distributed with a mean rate given by  $N_q = N_{tot}/4$ . Under this model, the likelihood of any given proposal for the models parameters  $(\alpha_c, \delta_c, \eta)$  will be

$$\mathcal{L} = p(N_1, N_2, N_3, N_4 | \alpha_c, \delta_c, \eta) = \mathcal{P}(N_1 | N_q) \cdot \mathcal{P}(N_2 | N_q) \cdot \mathcal{P}(N_3 | N_q) \cdot \mathcal{P}(N_4 | N_q) \quad (3.51)$$

where  $N_i, i = 1, 2, 3, 4$  is the number of sources in each quadrant, and  $\mathcal{P}(N_i | N_q)$  is the Poisson distribution with mean rate  $N_{tot}/4$  evaluated at  $N_i$ .

### 3.4.3 Models with biaxial symmetry

In this Section we extend the previous models to accommodate deviations from radial symmetry. We do this by allowing variations of the radial profile that depend on the angular coordinate but still maintain biaxial symmetry. This can be done in many ways. In this work we focus in the simplest one: the analytical expression of the radial profile is maintained along any radial direction but the profile parameters ( $r_c$  an  $r_t$  for example in the King profile) have an ellipse-like dependence on the angular coordinate.

This requires the definition of a coordinate system centred at the cluster centre and potentially rotated from the RA-Dec system of axes. We use the previously included the angle  $\eta$  between the principal axis of the ellipse and RA-Dec system. It is inferred together with the rest of the model parameters.

Starting with King's profile, the likelihood term now is defined as

$$p(R|r_c, r_t, k_1, \mathcal{M}_1) = k_1 \cdot R \cdot \left( \frac{1}{\sqrt{1 + (R/r_c(\phi))^2}} - \frac{1}{\sqrt{1 + (r_t(\phi)/r_c(\phi))^2}} \right)^2 \quad (3.52)$$

where

$$r_c(\phi) = \frac{r_{cx} \cdot r_{cy}}{\sqrt{((r_{cx} \sin(\phi))^2 + r_{cy} \cos(\phi))^2}}, \quad (3.53)$$

$$r_t(\phi) = \frac{r_{tx} \cdot r_{ty}}{\sqrt{((r_{tx} \sin(\phi))^2 + r_{ty} \cos(\phi))^2}}, \quad (3.54)$$

with  $r_{cx}$  and  $r_{cy}$  are the  $r_c$  ellipse semi-axes, and  $r_{tx}$  and  $r_{ty}$  are the  $r_t$  ellipse semi-axes. Notice that we do not constrain the two ellipses to have the same aspect ratio, but they are co-aligned.

The angle dependence of the OGKing profile is introduced according to

$$p(R|r_c, r_t, k_1, \mathcal{M}_1) = k_1 \cdot R \cdot \left( \left( 1 + (R/r_c(\phi))^{\frac{1}{0.4}} \right)^{-0.4} - \left( 1 + (r_t(\phi)/r_c(\phi))^{\frac{1}{0.4}} \right)^{-0.4} \right)^{1.2} \quad (3.55)$$

with  $r_c(\phi)$  and  $r_t(\phi)$  following the same angle dependence as in the classical King profile.

In the case of the RGP profile, the new likelihood is defined as

$$p(R|\alpha, \beta, \gamma = 0, R_b, k_3, \mathcal{M}_3) = \frac{k_3 \cdot R}{(1 + (R/R_b(\phi))^{1/\alpha})^{(-\beta)\alpha}}, \quad (3.56)$$

where the parameter  $R_b$  follows the same elliptic angle dependence as defined in Equations 3.53 and 3.54, and we do not incorporate any angle dependence for the exponents  $\alpha$  and  $\beta$ .

### 3.4.4 Models with luminosity segregation

In this Section, we consider an extension to introduce a dependence of the core radius with the  $J$  magnitude. We select the  $J$  magnitude because it is the reddest of the magnitudes that are available for all candidate members. We assume that stars of the same mass have approximately the same magnitude and that distance differences (due to the 3D spatial extend of the Pleiades) average out. The core radius dependence with the  $J$  magnitude is modelled as

$$r_c(\phi, J) = r_c(\phi) + \kappa \cdot (J - 13.5). \quad (3.57)$$

The slope of the relationship  $\kappa$  is independent of the position angle  $\phi$ . Therefore,

for  $J = 13.5$  (the mode of the  $J$  distribution) the model reduces to the elliptic profile described in Section 3.4.3. A positive value of  $\kappa$  corresponds to smaller values of the core radius for stars brighter than  $J = 13.5$ . In other words, it describes a system where the more massive stars are more concentrated than the less massive ones.

Finally, to compare the previous models, we use the model selection methodology described in Section 3.1.2 together with truncated uniform priors for all cluster parameters. The results of this analysis are detailed in Chapter 4.

## 3.5 Priors

The Bayesian formalism is characterised by the use of priors. These represent the objective way to establish the subjective beliefs that the user of the model may have about the distribution of the parameter values. Although the beliefs themselves remain subjective, the way to establish them is measurable and reproducible. For example, by stating that I use a normal distribution as the prior for certain parameter, I provide a measure of my subjective beliefs, its PDF, that others can reproduce.

In the following, I describe the information used to establish both the family of the prior distribution as well as its hyper-parameters (the parameters at the top hierarchy of the BHM). As mentioned before, these families are chosen to fall, whenever possible, in the category of weakly informative priors.

The priors in the BHM can be grouped into three main categories. The first one correspond to priors for parameters representing fractions in mixture models. The second and third categories correspond to parameters in the proper motions and photometric models.

In the BHM, there are different types of mixtures: the GMMs of the proper motions, the cluster-field mixture, and the singles-EMB mixture. At each mixture, the fractions quantify the contribution of each element in the mixture to the probability distribution. Fractions must add to one and be bounded by the  $[0, 1]$  interval. We choose the Dirichlet distribution to be the family of all fraction parameters. This decision roots in the fact that this distribution is the multivariate generalisation of the beta distribution. The latter is commonly used to model the probability of success of an event.

The Dirichlet distribution,  $Dir(\mathbf{x}|\boldsymbol{\alpha})$ , of dimension  $n$  ( $\{\mathbf{x}, \boldsymbol{\alpha}\} \subset \mathcal{R}^n$ ) has support in  $[0, 1]$  for each entry  $x_k$  of  $\mathbf{x}$ , and is parametrised by  $\boldsymbol{\alpha}$ . Each  $\alpha_k$  gives the concentration of the resulting PDF for entry  $x_k$ . The means and variances of these latters are given by,

$$E[x_k] = \frac{\alpha_k}{\sum_k \alpha_k}, \quad (3.58)$$

$$Var[x_k] = \frac{-\alpha_k \cdot (\alpha_k - \sum_k \alpha_k)}{(\sum_k \alpha_k)^2 \cdot (1 + \sum_k \alpha_k)}. \quad (3.59)$$

For the field-cluster mixture we set the hyper-parameters to  $\boldsymbol{\alpha} = \{98, 2\}$ . We expect a mean 98% of field objects and a 2% of cluster objects with little variance. These figures correspond to the existing prior knowledge, that is the fraction of field and cluster candidate members found by [Bouy et al. \(2015\)](#). For the single-EMB mixture we use an hyper-parameter value,  $\boldsymbol{\alpha}_{CB} = \{8, 2\}$ . We expect a mean 20% of EMB, as suggested by [Bouy et al. \(2015\)](#). For fractions in the proper motions GMM, hyper-parameter are  $\boldsymbol{\alpha}_{Cs} = \{1, 1, 5, 5\}$  and  $\boldsymbol{\alpha}_{Bs} = \{1.2, 8.8\}$ . These values induce fraction distributions whose means are similar to the fractions recovered after fitting a GMM to the candidate members of [Bouy et al. \(2015\)](#). For the fraction in the GMM of the *CI* distribution, the hyper-parameter were set all to 1, ( $\boldsymbol{\alpha}_{CI} = \{1, 1, 1, 1, 1\}$ ), which results in equal means and large variances for all **components in the mixture**.

**In the previous cases, with exception of the cluster-field mixture, the hyper parameters are chosen so that the resulting fraction distributions have large variances, see Fig.3.6.** The narrow variance in the cluster-field mixture expresses our prior belief about the number (fraction) of candidate members within our large data set.

**For the priors of the means in the proper motions GMM**, both of single stars and EMB, we choose the bivariate normal distribution. We set the hyper-parameters of this bivariate normal to those found after fitting a bivariate normal to the candidate members of [Bouy et al. \(2015\)](#). These values are

$$\boldsymbol{\mu}_{\mu_{pm}} = (16.30, -39.62) \text{ mas} \cdot \text{yr}^{-1},$$

and

$$\Sigma_{\mu_{pm}} = \begin{pmatrix} 36.84 & 1.18 \\ 1.18 & 40.71 \end{pmatrix} \text{ mas}^2 \cdot \text{yr}^{-2}.$$

As prior for the covariance matrices of both single stars and EMB proper motions we use the Half- $t(\nu, \mathbf{A})$  distribution. It is parametrised by a scalar  $\nu$  and a vector  $\mathbf{A}$ . As shown by [Huang & Wand \(2013\)](#), this distribution family leads to more accurate estimations of covariance matrices than the traditional Inverse-Wishart distribution. In specific, the marginal correlation parameters,  $\rho$ , have the following distribution,

$$p(\rho) \propto (1 - \rho^2)^{\frac{\nu}{2}-1}. \quad (3.60)$$

The standard deviation term  $\sigma_k$ , associated to entry  $k$ , is distributed according to Half- $t(\nu, A_k)$ . We set the hyper-parameters to  $\nu = 3$  and  $\mathbf{A}_{pm} = \{10^5, 10^5\}$  mas · yr<sup>-1</sup>. According to [Huang & Wand \(2013\)](#), arbitrarily large values of  $\mathbf{A}$  lead to arbitrarily weakly informative priors on the corresponding standard deviation terms.

Concerning the photometric priors, they can be grouped in three categories: (i) priors

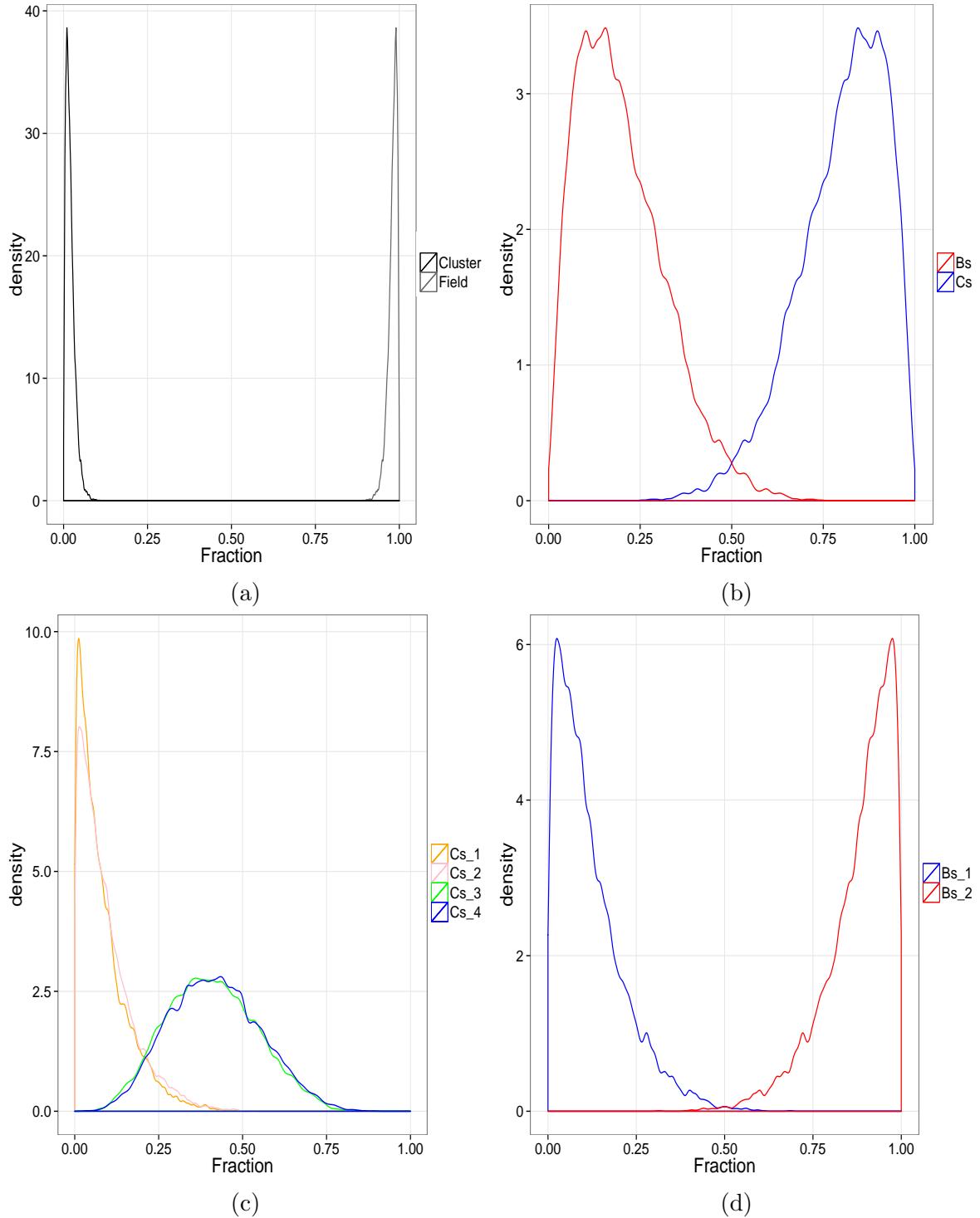


Figure 3.6: Kernel density of 10<sup>4</sup> realisations of the prior distributions of fraction parameters. Distributions of: (a) The field fraction  $\pi$ , (b) The EMB fraction,  $1 - \pi_{CB}$ , (c) the proper motions cluster fractions,  $\pi_{Cs}$ , and (d) the proper motions equal-mass binaries fractions  $\pi_{Bs}$ .

for the the *true CI*, (ii) priors for the splines coefficients, and (iii) priors for the cluster sequence intrinsic dispersion.

For the means in the univariate GMM of the *true CI*, I choose a uniform distribution in the range  $(0.8 \leq CI \leq 8)$ . For the standard deviations I choose the Half-Cauchy( $0, \eta$ )

distribution as suggested by [Gelman \(2006\)](#). The value of  $\eta$  is set to an arbitrarily large value,  $\eta = 100$ .

For the coefficients in the spline series we set the priors as univariate normal distributions. To find the mean and variance of these distributions we proceed as follows. First, we remove the EMB from the list of candidate members of [Bouy et al. \(2015\)](#). To do this, I performed an iterative fit of the cluster sequence, at each iteration I removed those objects whose photometry was as bright as that of the EMB. In the region of  $CI > 7$  there are no candidate members of [Bouy et al. \(2015\)](#) or of any other source. Thus, to provide a prior we complement our list of candidate members with the brown-dwarfs from the [Faherty et al. \(2012\)](#) sample. We choose only those objects observed in the same photometric bands of our data set. Finally, we fit the splines, and use the coefficients of this fit as the means,  $\mu_\beta$  of the univariate normal distributions. The standard deviation terms were set to  $\sigma_\beta = \{1, 1, 1, 1, 1, 0.5, 0.1\}$  mag. These values provide a reasonable compromise between cluster sequences compatible with the previously known candidates, and those far away or with exotic shapes. We show a sample of these priors in Fig. 3.7. This Figure shows also the brown-dwarfs from [Faherty et al. \(2012\)](#) and the sequence (dashed line) we use to provide the means of the univariate normal distributions.

Again, I choose again the Half-t( $\nu, \mathbf{A}$ ) distribution to set the prior for the parameters of the cluster intrinsic dispersion,  $\Sigma_{clus}$ . However, this time I use  $\mathbf{A}_{ph} = \{10, 10, 10, 10, 10\}$  mag. These values are large when compared to the standard deviation terms of the observed uncertainties. Therefore, they provide a weakly informative prior on the marginal standard deviation terms of the  $\Sigma_{clus}$  covariance matrix.

Table 3.2 shows a summary all the hyper-parameter values used for the prior distributions. Since the majority of these values come from the work of [Bouy et al. \(2015\)](#), it is reasonable to question the influence that this particular work has on the results of the present one.

The methodology of the Bayesian Hierarchical Models, the one chosen here, has been specifically designed to avoid as much as possible the subjectivity of the chosen priors ([Gelman 2006](#)). Thus, by using the results of [Bouy et al. \(2015\)](#) to set only the hyper-parameter values of the prior distributions we avoid much of the subjectivity of choosing that particular work. On top of that, the use of the weakly informative priors (see Section 3.2.1) further decreases the impact of the prior information on the recovered posterior distributions.

Despite the weakly informative priors and the BHM methodology, the work of [Bouy et al. \(2015\)](#) may have influence our results in a similar way as it was influenced by previous works. There is an unavoidable influence from the past into the present. In this work, we aim at keeping this influence at minimum by using the least subjective approaches to the inference process under a Bayesian framework.

Furthermore, as mentioned at the end of Section 3.2.1, the prior information must be analysed in terms of the posterior distribution and verify if they make sense (Gelman 2006; Gelman et al. 2013). Details of this analysis will be presented in Section 4.4.

The quantitative analysis of the influence that the work of Bouy et al. (2015) may have in the present one should wait until further evidence (the application of the present methodology to other star clusters) is collected.

Table 3.2: Hyper-parameters for different blocks of the Hierarchical Bayesian Model. The upper and middle blocks correspond, respectively, to hyper-parameters of the proper motions and photometric models. The lower block to hyper-parameters shared by both models.

Hp.	Value
$\alpha_{Cs}$	{1, 1, 5, 5}
$\alpha_{Bs}$	{1.2, 8.8}
$A_{pm}$	{ $10^5$ , $10^5$ }
$\mu_{\mu_{pm}}$	{16.30, -39.62}
$\Sigma_{\mu_{pm}}$	{36.84, 1.18, 40.71}
$\alpha_{CI}$	{1, 1, 1, 1, 1}
$rg_{CI}$	{0.8, 8}
$\eta$	100
$\mu_{\beta_Y}$	{7.65, 11.47, 10.66, 16.33, 16.49, 21.44, 22.49}
$\mu_{\beta_J}$	{7.61, 11.52, 10.20, 15.66, 15.58, 19.88, 21.16}
$\mu_{\beta_H}$	{7.63, 10.88, 9.50, 15.19, 15.04, 18.68, 20.64}
$\mu_{\beta_K}$	{7.55, 10.81, 9.32, 14.79, 14.62, 17.63, 20.24}
$\sigma_\beta$	{1, 1, 1, 1, 0.5, 0.1}
$A_{ph}$	{10, 10, 10, 10, 10}
$\alpha$	{98, 2}
$\alpha_{CB}$	{8, 2}
$\nu$	3

## 3.6 Sampling the posterior distribution

Theoretically, there are at least three possible approaches to obtain the posterior distributions of the parameters in our model. One of these options is the analytical approach. The computing of the posterior distribution will become intractable given the size of the data set and the high dimensionality of the parametric space; the BHM has 85 parameters. The second option is the use of a grid in the parametric space. The likelihood and the prior must be evaluated at each point in this

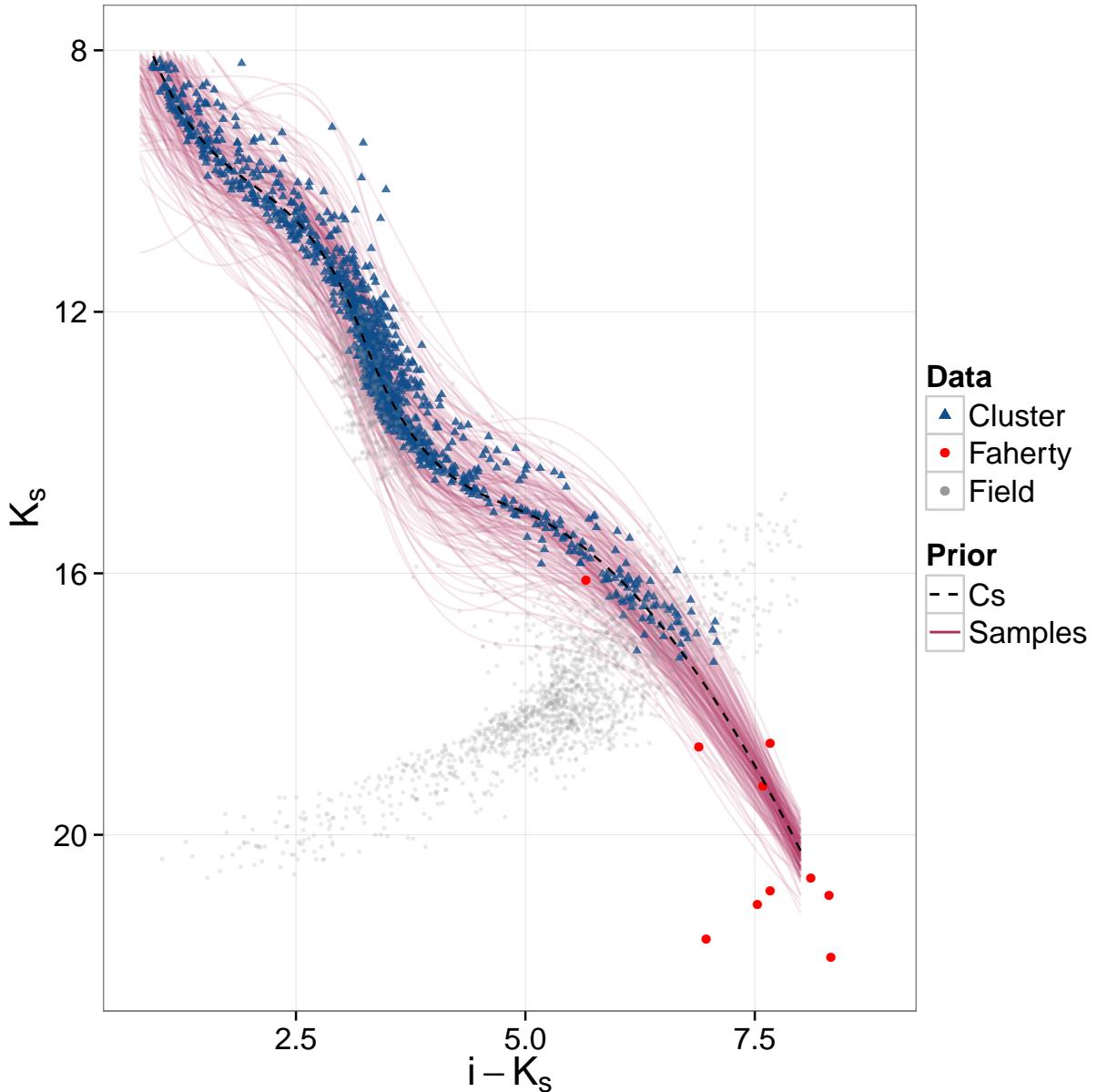


Figure 3.7: CMD  $K_s$  vs.  $i - K_s$  showing a sample (100 elements) of the prior for the coefficients in the splines series. Also shown are the brown-dwarfs from [Faherty et al. \(2012\)](#) sample (red dots), the cluster sequence (dashed line) used as mean for the priors, the candidate members of [Bouy et al. \(2015\)](#) (blue triangles), and a sample of the field ([grey dots](#)).

grid and then multiplied. This approach is reasonable when the parametric space is of moderate dimension ( $\leq 5$ ). It requires the evaluation of the posterior distribution  $q^p$  times, with  $q$  the number of grid points in one dimension, and  $p$  the dimension of the parametric space. The number of parameters in our model is 85, which immediately rules out this possibility. The third and so far only feasible approach is the use of Markov Chain Monte Carlo (MCMC) **sampling** methods. Although these methods provide a solution in a reasonable time, nevertheless, the bottle neck of computing time is due to the evaluation of the likelihood, which grows linearly with the size of the data set.

This Section is structured as follows. First, I introduce an heuristic technique to perform a fast search of the maximum a posteriori of our target distribution. Then, I describe the MCMC techniques available in the literature. **In particular, I focus on the one technique we choose, and the reasons of this decision.** The Section ends detailing the convergence assessment of the MCMC.

### 3.6.1 PSO

The likelihood of the data is the product of the individual likelihoods of each datum (Eq. 3.12). Therefore, the number of operations needed to evaluate the likelihood grows proportionally to the size of the data. As I will explain in Section 3.6.2, the burn-in phase of MCMC techniques allows them to reach the target distribution. However, once the MCMC reaches this target distribution, the burn-in computations are discarded. Since the evaluation of the likelihood, and therefore of the posterior, is computationally expensive, I decided to reduce as much as possible the burn-in phase. To do so, I provide MCMC with a set of initial solutions which are close to the Maximum-A-Posteriori (MAP) of the target (posterior) distribution. These near-MAP solutions must not be too crowded on the MAP, or otherwise the MCMC will spend too much time expanding them to reach the entire distribution. Here, there is a trade-off between the crowdedness of the solutions and its proximity to the MAP. Since we aim at obtaining a representative sample of the posterior distribution and not just an estimate of it, then the initial set of solution for the MCMC must be carefully chosen to minimise the computing time. This section provides the details of this procedure.

In Section 3.6.2, I will also show that the MCMC flavour more suitable to our objective belongs to the family of *ensemble* MCMC. This flavour works with particles in the parametric space. To make the transition between the initial near-MAP solutions and the MCMC particles as efficient as possible, I choose a MAP finder that works also with particles: the Particle Swarm Optimiser (PSO, [Kennedy & Eberhart 1995](#)). It provides a heuristic cheap-and-fast approach to the MAP solution. The PSO works with an ensemble of particles which move through the parametric space. These particles use the collective and individual past and present information to update their position. This information is specified by the score function, which in our case is the posterior distribution. The particles update their position iteratively according to their velocity. This velocity has a random but restricted magnitude. However, its direction is determined by the particle position, and the individual and collective positions with maximum score. [Kennedy & Eberhart \(1995\)](#) has a detailed description of the original algorithm, while a more efficient version is given by ([Clerc & Kennedy 2002](#)).

Although the PSO is a simple and rather efficient solution to the MAP approximation, it is far from perfect. Due to its heuristic origin, there is no theory behind its formulation.

Furthermore, it does not always guarantee the finding of the global maximum (for a convergence guaranteed version see [Patel et al. 2013](#)). Although, this issue does not affect our results (as we will see MCMC does guarantee the finding of the target distribution), it impacts the computing time. If the global maximum is not found in the PSO stage, then the MCMC will take longer to arrive to the target distribution.

On the other hand, the PSO stops its computations once the mean of the particles scores lies within a user defined tolerance. If this tolerance is too large, the PSO may stop far from the MAP. If it is too small, it may converge to the MAP but deliver solutions highly concentrated around it. This poses a problem to the following MCMC stage. For it to explore the full posterior distribution, the MCMC will need more iterations, thus more time, to expand the initially concentrated positions. **The optimal value for the relative tolerance is  $10^{-7}$ . I found it after several trials and errors. It was a time consuming exercise that will be avoided in the analyses of other clusters.**

To overcome the problem of crowdedness, I decide to use the charged PSO ([Blackwell & Bentley 2002](#)). Originally designed to optimise a time varying score function, the charged PSO maintains its exploratory capabilities due to an electrostatic force that repels particles when they get closer than a certain distance ([Blackwell & Bentley 2002](#)). Thanks to this electrostatic force the charged PSO avoids the over-crowding of particles around local best values.

The algorithm of [Blackwell & Bentley \(2002\)](#) computes distances in the entire parametric space. I find this approach unsuitable for our problem, thus I modified it. **This modified version and the values chosen for its parameters are described with more detail in Section 3.7, together with the rest of the developed codes.**

### 3.6.2 MCMC

#### Generalities

Markov Chain Monte Carlo (MCMC) is the generic name for a series of algorithms whose objective is the sampling of probability distributions. As their name indicates, the MCMC generates a chain (or a group of them) of Monte Carlo realisations that fulfil the Markov property. Monte Carlo realisations can be understood, broadly speaking, as continuous random realisations. Since it is an iterative algorithm, the chain is a process that refers to the joint of all random Monte Carlo steps. The Markov property indicates the probabilistic independence between steps in the chain that are separated more than one iteration. Thus, in a Markov chain, the probability of a future step depends only on the present step, and not in the past steps.

[Andrieu et al. \(2003\)](#) provides a brief and interesting summary of the history of the MCMC methods. In the following I use their work to describe the fundaments of MCMC. For more details, see the aforementioned authors and the book of [Brooks et al. \(2011\)](#).

A stochastic process is defined as a sequence  $\{\theta_1, \dots, \theta_n\}$  of random elements. On it, each element  $\theta_i \subset \mathbb{R}^k$ , with  $k$  the dimension of the *state space*.

A stochastic process,  $\boldsymbol{\theta} = \{\theta_0, \theta_1, \dots, \theta_n, \theta_{n+1}\}$  is called a Markov chain if

$$p(\theta_{n+1} | \theta_0, \theta_1, \dots, \theta_n) = p(\theta_{n+1} | \theta_n).$$

A Markov chain has two important distributions, the initial distribution and the transition distribution. The initial distribution is the marginal distribution of  $\theta_0$ ,  $p(\theta_0)$ . The transition distribution is the conditional probability  $p(\theta_{n+1} | \theta_n)$ . The latter is called stationary or homogeneous if it does not depend on  $n$ .

**If this transition is irreducible and aperiodic, then there is an *invariant* or *equilibrium distribution* to which the chain converges, regardless of the initial distribution. Here, aperiodic means that the chain does not have loops, while it is irreducible if the probability of exploring all other states is not zero.**

If we want to have  $p(\theta)$  as the invariant distribution, then it suffices that the transition distribution  $p_t(\cdot|\cdot)$  satisfies the detailed balance condition,

$$p(\theta_n) \cdot p_t(\theta_{n-1} | \theta_n) = p(\theta_{n-1}) \cdot p_t(\theta_n | \theta_{n-1}). \quad (3.61)$$

Thus, MCMC are Markov chains that satisfy the detailed balance condition, and have their invariant distribution as the target distribution. The large variety of MCMC algorithms arises from the efficiencies with which they arrive to the target distribution.

In the following I will review three of the MCMC categories: Metropolis-Hastings (MH), Hamiltonian Monte Carlo (HMC), and affine invariant samplers. The MH category comprises the classic MH algorithm but also contain particular cases like the Gibbs sampler ([Geman & Geman 1984](#)). I describe MH only for completeness and explanatory reasons. Later, I will focus on the particular cases of Hamiltonian Monte Carlo (HMC), and affine invariant for ensemble samplers. Finally, I will briefly describe Nested Sampling, an algorithm that uses MCMC to numerically compute the Bayesian evidence and simultaneously generate samples of the posterior distribution.

## Metropolis-Hastings

By far, the most popular MCMC algorithm is Metropolis-Hastings ([Metropolis et al. 1953](#); [Hastings 1970](#)). Once the Markov chain has been initialised in the state space, given the current  $\theta$  and the proposed  $\hat{\theta}$  positions, the chain moves from  $\theta$  to  $\hat{\theta}$  with acceptance probability:

$$\mathcal{A}(\hat{\theta} | \theta) = \min \left\{ 1, \frac{p(\hat{\theta}) \cdot q(\theta | \hat{\theta})}{p(\theta) \cdot q(\hat{\theta} | \theta)} \right\}, \quad (3.62)$$

where  $q$  is the transition probability. Since the algorithm allows rejection, it is aperiodic, and to ensure irreducibility, the support of  $q$  must include that of  $p$  (Andrieu et al. 2003). The popularity of MH lies in its simplicity. Nevertheless it requires a careful tuning of the transition probability. Usually, this probability is given by a normal distribution. It works well for relatively low dimensions of the parametric space ( $\leq 5$ ). However, once the dimension goes higher, the MH algorithm spends a great amount of time tuning the parameters of this multivariate normal distribution. In particular, those of its covariance matrix.

### Hamiltonian Monte Carlo

The Hamiltonian Monte Carlo algorithms (Duane et al. 1987; Neal 1996), as their name suggest<sup>13</sup>, use Hamiltonian dynamics to express the target distribution as the potential distribution of a hamiltonian system of particles. In such systems the total energy is the sum of the potential and kinetic energies. The potential distribution depends only on position, whereas the kinetic one on momentum. HMC introduces a momentum to the particles in order to use their positions as a sample of the target distribution. To update the particles positions, HMC uses the Hamilton equations, which contain information about the gradient of the potential. Once HMC has tuned the momentum distribution, the proposed positions are more likely in terms of the target distribution. Therefore, using the information about the gradient of the target distribution, HMC is able to improve the acceptance ratio of the proposed steps. A detailed description of HMC can be found in Chapter 5 of Brooks et al. (2011). The package *Stan* (Carpenter et al. 2017) provides an efficient implementation of HMC.

### Affine invariant

Affine invariant MCMC samplers use many particles, the ensemble, to sample the target distribution with a performance that is independent of its shape in the parametric space. Affine invariant MCMC does not need the tuning the transition probability. For this reason, these samplers are faster than standard MCMC (Goodman & Weare 2010). In the following I use the derivation of Goodman & Weare (2010).

An ensemble  $\theta$  is a set of  $L$  particles  $\theta_l \subset \mathbb{R}^k$ . It lives in state space  $\mathbb{R}^{kL}$ , and the positions of it particles are independently drawn from the target distribution  $\pi$ . Therefore,

$$\Pi(\theta) = \pi(\theta_1) \cdot \pi(\theta_2) \dots \pi(\theta_L).$$

Thus, an ensemble MCMC is a Markov chain in the state space of ensembles. An ensemble MCMC preserves the equilibrium distribution without the individual particles

---

<sup>13</sup>Originally called Hybrid Monte Carlo by (Duane et al. 1987)

sequence,  $\theta_1(1), \theta_1(2), \dots \theta_1(t)$ , being Markov or even independent. However, to update the particles positions, the detailed balance condition (Eq. 3.61) must be fulfilled. [Goodman & Weare \(2010\)](#) use partial resampling to ensure this. In partial resampling, the transition probability preserves the target (invariant) distribution if the single particle steps preserve the conditional distribution of the particle position given the complementary ensemble (the rest of the particles). Using the affine invariant *stretch move* (see below), these authors are able to define a Markov chain, in the state space of ensembles, that satisfies the detailed balance condition.

The stretch move  $\theta_k(t) \rightarrow \hat{\theta}$  is defined as,

$$\hat{\theta} = \theta_j(t) + z \cdot (\theta_k(t) - \theta_j(t)),$$

where  $\theta_j(t)$  is the current position of a particle in the complementary ensemble, and  $z$  is the stretching factor. It produces a symmetric transition,  $p(\theta_k(t) \rightarrow \hat{\theta}) = p(\theta_k(t) \leftarrow \hat{\theta})$ , if its density  $g(z)$  satisfies the symmetry condition

$$g\left(\frac{1}{z}\right) = z \cdot g(z).$$

Finally, [Goodman & Weare \(2010\)](#) define their affine invariant MCMC using the following distribution for  $g(z)$ ,

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{for } z \in [1/a, a] \\ 0 & \text{for } z \notin [1/a, a] \end{cases} \quad (3.63)$$

and the acceptance probability,

$$\mathcal{A}(\hat{\theta}|\theta) = \min \left\{ 1, z^{n-1} \cdot \frac{p(\hat{\theta})}{p(\theta)} \right\}. \quad (3.64)$$

The parameter  $a$ , which must be greater than one, improves the performance of the sampler ([Goodman & Weare 2010](#)). The acceptance fraction of the proposed transitions depends both on the ratio of probabilities  $p(\hat{\theta})/p(\theta)$  and on the value of  $z^{n-1}$ . The support of the latter depends on the parameter  $a$ . Given  $g(z)$  and certain dimension of the parametric space, increasing  $a$  results in more probable smaller values of  $z$ , thus in smaller acceptance probabilities.

One of the great advantages of MCMC ensemble samplers is its possibility of parallelisation. Since they work with particles, these particles can be distributed among cores in a computer cluster, therefore reducing the computing time when compared to non ensemble MCMC. [Foreman-Mackey et al. \(2013\)](#) implemented the affine invariant stretch move of [Goodman & Weare \(2010\)](#) in the Python package *emcee*.

### Nested sampling

Nested sampling ([Skilling 2004, 2006](#)) is an algorithm designed to numerically integrate the evidence (Eq. 3.15). As a by-product, it also delivers a sample of the posterior distribution.

To compute the evidence integral, it uses  $N$  particles whose positions, in the parametric space, are sampled from the prior. Then, at each subsequent step  $i$  the algorithm computes the likelihoods of the  $N$  particles. The particle with lowest likelihood is stored as  $x_i$  together with its likelihood  $L_i$ . The weight  $w_i$  is computed as

$$w_i = e^{-\frac{(i-1)}{N}} - e^{-\frac{i}{N}}.$$

The particle  $x_i$  is replaced with a new draw from the prior, on the condition that its likelihood is greater than  $L_i$ .

Once certain number of iterations have been done, the evidence integral is approximated by

$$z \leftarrow \sum_i w_i \cdot L_i. \quad (3.65)$$

The original algorithm was designed to compute the evidence of a unimodal distribution. However, an improved version of the original algorithm was implemented in *MultiNest* ([Feroz et al. 2009](#)). This version allows the sampling and computing of evidence in the even more difficult multimodal posteriors.

### 3.6.3 Implementation and convergence: PSO and MCMC

To sample the posterior distribution in our problem, we choose *emcee* due to the following properties: i) the affine invariance allows a faster convergence over common and skewed distributions (see [Goodman & Weare 2010; Foreman-Mackey et al. 2013](#), for details), ii) it can be run in parallel by distributing particles over the nodes of a computer cluster, which reduces considerably the computing time; and iii) it requires the hand-tuning of only two constants: the number of particles, and the parameter  $a$  of the  $g(z)$  distribution(Eq. 3.63). I choose a ratio of particles to parameters of two, that results in 170 particles. This is the minimum ratio recommended by [Foreman-Mackey et al. \(2013\)](#), which still allows a reasonable computing time. After trial and error, I fix the value of the  $a$  parameter to  $a = 1.3$ . As mentioned by [Goodman & Weare \(2010\)](#), this parameter can be tuned to improve performance of the sampler. This value keeps the acceptance fraction in the range  $0.2 - 0.5$ , as recommended by [Foreman-Mackey et al. \(2013\)](#).

As a front-end of *emcee*, and to handle the input and output of data, I use a modified version of the *emcee* handler known as *Cosmo Hammer* ([Akeret et al. 2013](#)). The next Section provides the details of these modifications.

As mentioned earlier, the PSO does not guarantee the finding of the global maximum of the score function. Therefore, I implement an iterative approach that minimises the risk of the PSO getting stuck in local maxima. To do so, I iteratively run PSO and 50 iterations of *emcee* (with the same number of particles as the PSO) until the relative difference between means of consecutive iterations is lower than  $10^{-7}$ . The iterations of *emcee* spread the PSO solution without moving away from the target distribution.

Neither scheme, PSO alone or PSO-*emcee*, guarantees finding the global maximum. The solution this scheme provides could indeed be biased. However, we use them to obtain only a fast estimate of the global maximum, or at least, of points in its vicinity. If the initial solution provided by this scheme is indeed biased, the final *emcee* run erases, during the burning phase, any dependance on the initial solutions. After the convergence of the PSO-*emcee* scheme, I run *emcee* alone until it converges.

Convergence to the target distribution occurs when each parameter enters into the stationary equilibrium or normal state. The Central Limit Theorem ensures that this state exists. See [Roberts & Rosenthal \(2004\)](#) for guaranteeing conditions and [Goodman & Weare \(2010\)](#) for *irreducibility* of the *emcee* stretch move. The stationary or normal state is reached when, in at least 95% of the iterations, the sample mean is bounded by two standard deviations of the sample, and the variance by the two standard deviation of the variance <sup>14</sup>. Fig. 3.8 shows the mean and variance of the ensemble of *emcee* particles for the last 4000 iterations before stoping the sampling. Both the mean and variances are normalised to the value of the last iteration.

---

<sup>14</sup>  $sd(\sigma^2) = \sigma^2 \sqrt{\kappa/n + 2/(n-1)}$  with  $\kappa$  the kurtosis and  $n$  the sample size.

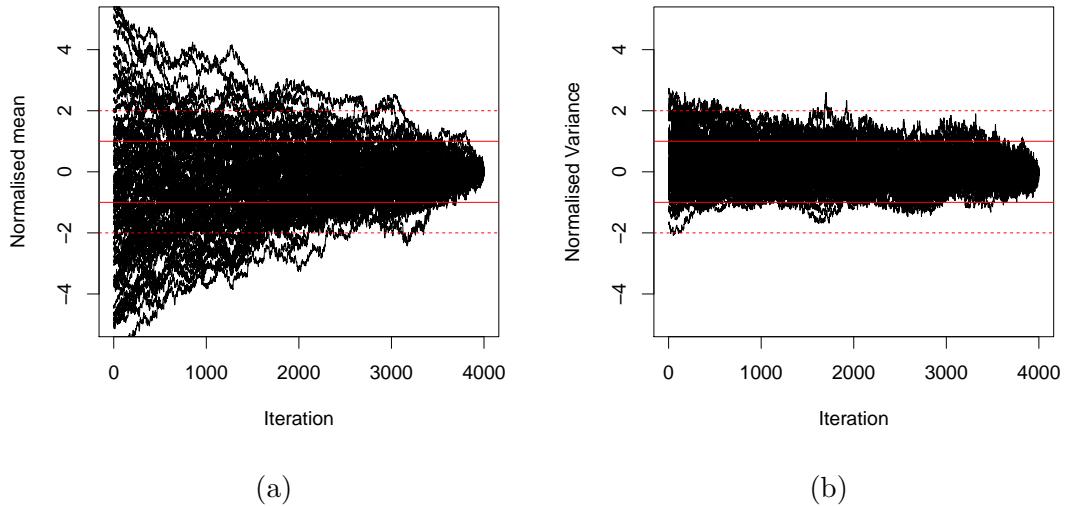


Figure 3.8: Normalised mean (a) and variance (b) of each parameter in BHM model as functions of iterations. The normalisation values are the mean and variance of the ensemble of particles positions at the last iteration. Red lines show one and two sigma levels of these normalisation values. Only shown the last 4000 iterations previous to stopping the algorithm.

I stop the *emcee* sampling once all parameters have entered the equilibrium state and the criterium of [Gong & Flegal \(2016\)](#)<sup>15</sup> is fulfilled. We choose this criterium because it was developed for high-dimensional problems and tested on Hierarchical Bayesian Models, as in the present work. In this criterium, the MCMC chain stops once its effective sample size (ESS) is larger than a minimum sample size. This minimum is computed using the required accuracy,  $v$ , for each parameter confidence interval  $(1 - \delta) \cdot 100\%$ . The ESS is the size that an independent and identically distributed sample must have to provide the desired accuracy on the parametric inference.

The *emcee* run stops once the ESS of the ensemble of walkers is greater than the minimum sample size needed for the required accuracy  $\epsilon = 0.05$  on the 68% confidence interval ( $\delta = 0.32$ ) of each parameter.

## 3.7 Codes

This Section sets out the details about the code I developed to perform the computation described throughout this chapter. First, I give a brief chronological description of the model development. Later, I will describe the details on the implementation of the charged PSO, the modified *emcee*, and the GMM used to describe the field population. Finally, I will end this Section detailing the hybrid high performance computing (Hybrid-HPC) code developed to minimise the computing time of the posterior distribution in the BHM.

<sup>15</sup> Implemented in the R package *mcmcse* ([Flegal et al. 2016](#))

The first version of the Bayesian Hierarchical Model was implemented by Ángel Berihuete in the package *Stan* (Carpenter et al. 2017). It comprised a Bayesian model of the ML model of Sarro et al. (2014). The proper motions were modelled using a single mixture of gaussians. The photometry was modelled with a Chebyshev polynomial parametrised by the length along the sequence. This length was found using a principal curve analysis and lacked physical interpretation.

I took this version and modified it in the following aspects. I included the photometric and proper motions EMB sequence, the uncertainties both in proper motion and in the photometry, and the width of the sequence modelled as the multivariate gaussian. Then, we realised that the principal curve analysis is not compatible with the deconvolution methodology. **The principal curve analysis finds the dominant curve in the observed data, and not the *true* underlying relation that generates the observed data once the individual noise process for each object is accounted for.** For this reason, the principal curve analysis is affected by individual uncertainties (see Hong et al. 2016, for the negative impact of heteroscedastic data on the related principal component analysis). Instead, we decided to model the intrinsic *true* underlying photometric relation with polynomials. We use as parameter for these polynomials the *true* colour  $CI$ , which is a more interpretable parameter than the distance along the principal curve, which was used in the previous version. Since we have one *true*  $CI$  for each object, the number of parameters is equal to the number of objects. We marginalised all these nuisance parameters with the aid of a prior. For this prior we introduced a probability distribution modelled by a GMM.

Previous to the introduction of the marginalisation of the nuisance parameters, the model worked fine on samples of a few hundreds of stars. Once the marginalisation was introduced, the computing time of the model increased dramatically, rendering its application to higher data sizes impractical. At this point we decided to port the existing *Stan* code into *Python*<sup>16</sup> so that we could work with the parallel *emcee* code. *emcee* proved to be of great use. Due to its parallelisation capabilities we were able to increase the data size from 2000 to 10,000 objects. Since the computing of the likelihood was the highest computational challenge, I developed my own routines to perform it in parallel. However, *CosmoHammer* (Akeret et al. 2013) turned out to be more efficient in distributing the parallel loads. I ported the BHM code into *CosmoHammer* and modified the latter. The modifications ranged from data files and log entries to the introduction of priors and the handling of a Hybrid-HPC scheme using both MPI and multithreading. Despite the Hybrid-HPC scheme, the computing of the likelihood of a data set with  $10^5$  objects seemed unreachable. At this point, I performed two tasks: the first was to strip the code of all auxiliary libraries calls, and the second was the vectorisation of the majority of the

---

<sup>16</sup><https://www.python.org>

operations. Since the parameters of the field were held fixed, the field likelihood was computed externally for each object. The code was then fed with the data set, the field likelihood, and all the auxiliary computations reduced to a minimum. Among the reduced computation there are, for example, the Cholesky decompositions and matrix inversions of the covariance matrices of the uncertainty. Instead of doing these computation inside the code, the code was fed with the precomputed values. This further reduced the computing time.

Introducing PSO and later the charged PSO further reduced the computing time. At this point the code was able to run on a data set with  $10^5$  objects. However, convergence of the MCMC still required several weeks of computations. Once the approximation to the marginalisation integral (Eqs. 3.33 and 3.34) was introduced, the computing time reduced far more. Finally, the tuning of the *emcee* parameters allowed us to increase the acceptance fraction, and reach convergence within four weeks of full computing time with an 80 cores computer cluster. It is indeed a very long time. However it is reasonable compared with our original estimates of approximately 2 years of computing time<sup>17</sup>.

### 3.7.1 The modified charged PSO

As explained before, the charged PSO of Blackwell & Bentley (2002) was inappropriate to our objective. The metric of the parametric space of our problem is not isotropic because parameters have different length scales. For example, while fractions are constrained in the  $[0, 1]$  interval, proper motions parameters are allowed in the range of proper motion measurements  $[-99, 99] \text{ mas} \cdot \text{yr}^{-1}$ . Therefore, the use of an isotropic metric results in a solution which is crowded in some parameters while is over-dispersed in others. To solve this issue, I modified the charged PSO by measuring distance between particles and applying the electrostatic force independently on each parameter. In such a way, the electrostatic force plays a role only when the relative distance between particles in any given parameter is smaller than  $10^{-10}$ . I found this value heuristically.

In the original version of Blackwell & Bentley (2002), each particle is subject to the acceleration,

$$\mathbf{a} = \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}^3} \cdot \mathbf{r}_{ij}, \quad p_{\text{core}} < r_{ij} < p \quad (3.66)$$

where  $q_i$  and  $q_j$  are the charges of particles  $i$  and  $j$ , and  $r_{ij}$  is the distance between them. The distances  $p_{\text{core}}$  and  $p$  indicate the minimum and maximum distances at which the electrostatic force comes into action. Outside this range, the electrostatic force is zero. In this equation,  $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ , where  $\mathbf{x}_i, \mathbf{x}_j$  are the positions of particles  $i$  and  $j$ . Also,  $\mathbf{r}_{ij}, \mathbf{x}_i, \mathbf{x}_j \subset \mathbb{R}^d$ , with  $d$  the dimension of the space.

In the modified version, the distance is measured independently in each dimension

---

<sup>17</sup>Today, the DANCe team is working on a GPU version of the code which computes the same amount of calculations in a couple of days.

of the parametric space. Thus,  $\mathbf{r}_{ij} = \{x_{1,i} - x_{1,j}, x_{2,i} - x_{2,j}, \dots, x_{d,i} - x_{d,j}\}$ . Also the acceleration has the form,

$$\mathbf{a} = \sum_{i \neq j} \frac{\mathbf{q}_i \cdot \mathbf{q}_j}{r_{ij}^2} \cdot \mathbf{r}_{ij}, \quad 10^{-50} < \frac{r_{ij}}{r_{eq}} < \epsilon \quad (3.67)$$

and it is now applied over each dimension of the parametric space. The distance  $r_{eq}$  is that at which the velocity caused by the acceleration equals the mean velocity caused by the common PSO.  $\epsilon$  is a free parameter which, as said previously, was set heuristically to  $10^{-10}$ .

### 3.7.2 Improvements of emcee

The modification I introduced in *emcee*, although very simple, improved the acceptance fraction and mixing of the particles. To allow the parallelisation, [Foreman-Mackey et al. \(2013\)](#) divide the ensemble of particles in two ensembles. In the original version, the particles in one ensemble use one and the same particle in the complementary ensemble to compute their positions according to Eq. 3.63. In the modified version, particles from one ensemble update their positions using a particle from the complementary ensemble. However, this particle is chosen randomly at each iteration.

In a private communication with David Foreman, the developer of *emcee*, he mentions that a similar modification was already introduced in a beta version of the *emcee* code.

### 3.7.3 GMM for the field population

As mentioned earlier in this Chapter, the field population is modelled by means of two independent photometric and proper motion distributions. The MLE of the parameters of these distributions were found using the EM algorithm. The conventional EM algorithm for GMM ([Dempster et al. 1977](#)) for a mixture of  $M$  gaussians goes as follows. Given a set of parameters  $\theta = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^M$ , where  $w_i$ ,  $\boldsymbol{\mu}_i$ , and  $\boldsymbol{\Sigma}_i$  are the fraction, mean and covariance matrix of gaussian component  $i$ , the likelihood of the data is,

$$p(\{\mathbf{y}_n\}_{n=1}^N | \theta) = \prod_{n=1}^N \sum_{i=1}^M w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (3.68)$$

To solve the problem, the EM algorithm requires a set of  $N$  variables,  $\{\mathbf{z}_n\}_{n=1}^N$ , of dimension  $M$ . The variable  $z_{n,i}$  represent the probability that observation  $y_n$  was drawn from gaussian  $i$ . Therefore,

$$1 = \sum_{i=1}^M z_{n,i}. \quad (3.69)$$

These  $\mathbf{z}$  latent variables are found as

$$z_{n,i} = \frac{w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{i=1}^M w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}. \quad (3.70)$$

The EM works, as its name indicates, by maximising the expected value of the likelihood. The latter is given by

$$E[p(\{\mathbf{y}_n\}_{n=1}^N | \theta)] = \prod_{n=1}^N \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (3.71)$$

The previous expectation is maximal when,

$$w_i = \frac{1}{N} \sum_{n=1}^N z_{n,i}, \quad (3.72)$$

$$\boldsymbol{\mu}_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \sum_{n=1}^N z_{n,i} \cdot \mathbf{y}_n, \quad (3.73)$$

$$\boldsymbol{\Sigma}_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \sum_{n=1}^N z_{n,i} \cdot (\mathbf{y}_n - \boldsymbol{\mu}_i) \times (\mathbf{y}_n - \boldsymbol{\mu}_i)^T. \quad (3.74)$$

The modified version of the GMM, which includes a uniform distribution, is now a particular case of the GMM. This can be viewed as a gaussian distribution with fixed parameters and a constant probability  $c$  given by the uniform distribution. The new expectation is then,

$$E[p(\{\mathbf{y}_n\}_{n=1}^N | \theta)] = \prod_{n=1}^N \left[ z_{n,0} \cdot w_0 \cdot c + \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right]. \quad (3.75)$$

The maximisation step remains identical except for the indices. There are now  $M + 1$  fractions  $w_i$ , with  $i = 0, 1, \dots, M$ , and the means and covariances run from  $i = 1, \dots, M$ .

Regarding the photometric GMM of the field, and given that the photometry has missing values, we use the EM algorithm of [McMichael \(1996\)](#). This algorithm was developed to obtain the MLE of data sets containing objects with missing values. A more recent version has been developed by [Lin et al. \(2006\)](#), however, this assumes that the missing entries are randomly distributed (missing at random). However, as mentioned earlier, the missing values in the photometry of the DANCE DR2 are not randomly distributed.

In the algorithm of [McMichael \(1996\)](#), there is a set of  $N$  gain matrices, one for each datum. Each  $M_n$  matrix is an identity matrix in which the rows of the corresponding missing value have been deleted. Thus, the expected value of the likelihood is now,

$$E[p(\{\mathbf{y}_n\}_{n=1}^N | \theta)] = \prod_{n=1}^N \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | M_n \boldsymbol{\mu}_i, M_n \boldsymbol{\Sigma}_i M_n). \quad (3.76)$$

The maximisation step is now,

$$w_i = \frac{1}{N} \sum_{n=1}^N z_{n,i}, \quad (3.77)$$

$$\boldsymbol{\mu}_i = \frac{\sum_{n=1}^N z_{n,i} \cdot H_n \mathbf{y}_n}{\sum_{n=1}^N z_{n,i} H_n M_n} \quad (3.78)$$

with

$$H_i = M_i^T (M \boldsymbol{\Sigma}_i M^T)^{-1} \quad (3.79)$$

The maximisation step has no analytical solution for the covariance matrix. Therefore, a modified steepest descent is used

$$\boldsymbol{\Sigma}_i \leftarrow \boldsymbol{\Sigma}_i + \frac{\rho}{2} \cdot \boldsymbol{\Sigma}_i \Delta_i \boldsymbol{\Sigma}_i, \quad (3.80)$$

with  $\Delta_i$  given by

$$\Delta_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \cdot \sum_{n=1}^N z_{n,i} \cdot [H_n (\mathbf{y}_n - M_n \boldsymbol{\mu}_i) \times (\mathbf{y}_n - M_n \boldsymbol{\mu}_i)^T H_n^T - H_n I]. \quad (3.81)$$

This algorithm preserves the monotonic convergence of the conventional EM, and returns positive definite matrices provided that  $\rho < 2$ . For details and validation of this algorithm see [McMichael \(1996\)](#).

### 3.7.4 Hybrid-HPC implementations

As I outlined before, the parallel computing approach was an unavoidable step. Once the code was ported to *Python* and *CosmoHammer*, I modified the latter to better fit our needs. The modifications were mainly on the management of input and output files and the python *multiprocessing* package for the multithreaded computing of likelihoods. Also, I striped some of its original functions to reduce memory usage and implemented some others, like the use of initial positions for the particles.

In the Hybrid-HPC approach, the particles of *emcee* are distributed on the nodes of the computing cluster by means of the MPI protocol. Then, each core in the node computes the likelihood of one fraction of the objects in the data set. This Hybrid-HPC code was implemented and tested in different computing cluster architectures. For the cluster at the Centre of Astrobiology (CAB, Villanueva de la cañada, Madrid, Spain), I

used a configuration of 6 nodes each with 12 cores. For the cluster at the University of Cádiz, (Andalucía, Spain) I used a configuration of 5 nodes each with 16 cores.

However, the Hybrid-HPC approach was not the best solution at the Infrastructure de Calcul Intensive et de Donées<sup>18</sup> of the University of Grenoble Alpes. At the Froggy<sup>19</sup> cluster, the code continuously render errors of communication. For this reason, I implemented a MPI-only version of the code. In this version, the multithreading approach is left aside. Instead, the totality of the available cores is fully dedicated to the computing of the likelihood. Each of the  $n$  cores computes the likelihood of the  $n$ th fraction of the objects in the data set. Once the likelihood of all particles has been computed, the master node evaluates the new positions of the particles.

I finish this chapter with a brief description of the difficulties faced in the development and testing of the BHM code. As the code evolved in complexity, my computational skills were compelled to evolve as well. I started by learning R and solving some toy problems on it. Later, when the dimensionality of the posterior increased I learned *Stan*. When the data set increased in size, we faced the parallelisation, so I learned Python and MPI. Once these versions were operable, I was forced to deal with libraries, from the common, numpy, scipy and numba, to the linking of modules and libraries. Finally, when dealing with several clusters I learned the queue languages Condor, slurm and OAR. Currently, I am working on the improvement (memory allocation and data distribution) of the GPU implementation of the BHM code.

---

<sup>18</sup><https://gricad.univ-grenoble-alpes.fr>

<sup>19</sup><https://ciment.ujf-grenoble.fr/wiki-pub/index.php/Hardware:Froggy>

Table 3.1: Parameters names, symbols, and priors.

Name	Symbol	Prior	Name	Symbol	Prior
Field fraction	$\pi$	Dirichlet( $\alpha$ )	Coefficient [1,1]	$\beta_{Y,1}$	Normal( $\mu_\beta, \sigma_\beta$ )
Cs fraction	$\pi_{CB}$	Dirichlet( $\alpha_{Cs}$ )	Coefficient [1,2]	$\beta_{Y,2}$	Normal( $\mu_\beta, \sigma_\beta$ )
Cs PM fraction 1	$\pi_{Cs,1}$	Dirichlet( $\alpha_{Cs}$ )	Coefficient [1,3]	$\beta_{Y,3}$	Normal( $\mu_\beta, \sigma_\beta$ )
Cs PM fraction 2	$\pi_{Cs,2}$	Dirichlet( $\alpha_{Cs}$ )	Coefficient [1,4]	$\beta_{Y,4}$	Normal( $\mu_\beta, \sigma_\beta$ )
Cs PM fraction 3	$\pi_{Cs,3}$	Dirichlet( $\alpha_{Cs}$ )	Coefficient [1,5]	$\beta_{Y,5}$	Normal( $\mu_\beta, \sigma_\beta$ )
Bs PM fraction 1	$\pi_{Bs,1}$	Dirichlet( $\alpha_{Bs}$ )	Coefficient [1,6]	$\beta_{Y,6}$	Normal( $\mu_\beta, \sigma_\beta$ )
Color fraction 1	$\pi_{CI,1}$	Dirichlet( $\alpha_{CI}$ )	Coefficient [1,7]	$\beta_{Y,7}$	Normal( $\mu_\beta, \sigma_\beta$ )
Color fraction 2	$\pi_{CI,2}$	Dirichlet( $\alpha_{CI}$ )	Coefficient [2,1]	$\beta_{J,1}$	Normal( $\mu_\beta, \sigma_\beta$ )
Color fraction 3	$\pi_{CI,3}$	Dirichlet( $\alpha_{CI}$ )	Coefficient [2,2]	$\beta_{J,2}$	Normal( $\mu_\beta, \sigma_\beta$ )
Color fraction 4	$\pi_{CI,4}$	Dirichlet( $\alpha_{CI}$ )	Coefficient [2,3]	$\beta_{J,3}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean color 1	$\mu_{CI,1}$	Unifrom( $rg_{CI}$ )	Coefficient [2,4]	$\beta_{J,4}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean color 2	$\mu_{CI,2}$	Unifrom( $rg_{CI}$ )	Coefficient [2,5]	$\beta_{J,5}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean color 3	$\mu_{CI,3}$	Unifrom( $rg_{CI}$ )	Coefficient [2,6]	$\beta_{J,6}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean color 4	$\mu_{CI,4}$	Unifrom( $rg_{CI}$ )	Coefficient [2,7]	$\beta_{J,7}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean color 5	$\mu_{CI,5}$	Unifrom( $rg_{CI}$ )	Coefficient [3,1]	$\beta_{H,1}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance color 1	$\sigma_{CI,1}$	HalfCauchy(0, $\eta$ )	Coefficient [3,2]	$\beta_{H,2}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance color 2	$\sigma_{CI,2}$	HalfCauchy(0, $\eta$ )	Coefficient [3,3]	$\beta_{H,3}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance color 3	$\sigma_{CI,3}$	HalfCauchy(0, $\eta$ )	Coefficient [3,4]	$\beta_{H,4}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance color 4	$\sigma_{CI,4}$	HalfCauchy(0, $\eta$ )	Coefficient [3,5]	$\beta_{H,5}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance color 5	$\sigma_{CI,5}$	HalfCauchy(0, $\eta$ )	Coefficient [3,6]	$\beta_{H,6}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean PM Cs[1,1]	$\mu_{Cs,1}$	Normal( $\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$ )	Coefficient [3,7]	$\beta_{H,7}$	Normal( $\mu_\beta, \sigma_\beta$ )
Mean PM Cs[1,2]	$\mu_{Cs,2}$	Normal( $\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$ )	Coefficient [4,1]	$\beta_{K,1}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[1,1]	$\Sigma_{Cs,1,1}$	Half-t( $\nu, A_{pm}$ )	Coefficient [4,2]	$\beta_{K,2}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[1,2]	$\Sigma_{Cs,1,2}$	Half-t( $\nu, A_{pm}$ )	Coefficient [4,3]	$\beta_{K,3}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[1,3]	$\Sigma_{Cs,1,3}$	Half-t( $\nu, A_{pm}$ )	Coefficient [4,4]	$\beta_{K,4}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[2,1]	$\Sigma_{Cs,2,1}$	Half-t( $\nu, A_{pm}$ )	Coefficient [4,5]	$\beta_{K,5}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[2,2]	$\Sigma_{Cs,2,2}$	Half-t( $\nu, A_{pm}$ )	Coefficient [4,6]	$\beta_{K,6}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[2,3]	$\Sigma_{Cs,2,3}$	Half-t( $\nu, A_{pm}$ )	Coefficient [4,7]	$\beta_{K,7}$	Normal( $\mu_\beta, \sigma_\beta$ )
Variance Cs[3,1]	$\Sigma_{Cs,3,1}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[1]	$\Sigma_{clus}[1]$	Half-t( $\nu, A_{ph}$ )
Variance Cs[3,2]	$\Sigma_{Cs,3,2}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[2]	$\Sigma_{clus}[2]$	Half-t( $\nu, A_{ph}$ )
Variance Cs[3,3]	$\Sigma_{Cs,3,3}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[3]	$\Sigma_{clus}[3]$	Half-t( $\nu, A_{ph}$ )
Variance Cs[4,1]	$\Sigma_{Cs,4,1}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[4]	$\Sigma_{clus}[4]$	Half-t( $\nu, A_{ph}$ )
Variance Cs[4,2]	$\Sigma_{Cs,4,2}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[5]	$\Sigma_{clus}[5]$	Half-t( $\nu, A_{ph}$ )
Variance Cs[4,3]	$\Sigma_{Cs,4,3}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[6]	$\Sigma_{clus}[6]$	Half-t( $\nu, A_{ph}$ )
Mean PM Bs[1,1]	$\mu_{Bs,1}$	Normal( $\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$ )	Covariance Phot[7]	$\Sigma_{clus}[7]$	Half-t( $\nu, A_{ph}$ )
Mean PM Bs[1,2]	$\mu_{Bs,2}$	Normal( $\mu_{\mu_{pm}}, \Sigma_{\mu_{pm}}$ )	Covariance Phot[8]	$\Sigma_{clus}[8]$	Half-t( $\nu, A_{ph}$ )
Variance Bs[1,1]	$\Sigma_{Bs,1,1}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[9]	$\Sigma_{clus}[9]$	Half-t( $\nu, A_{ph}$ )
Variance Bs[1,2]	$\Sigma_{Bs,1,2}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[10]	$\Sigma_{clus}[10]$	Half-t( $\nu, A_{ph}$ )
Variance Bs[1,3]	$\Sigma_{Bs,1,3}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[11]	$\Sigma_{clus}[11]$	Half-t( $\nu, A_{ph}$ )
Variance Bs[2,1]	$\Sigma_{Bs,2,1}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[12]	$\Sigma_{clus}[12]$	Half-t( $\nu, A_{ph}$ )
Variance Bs[2,2]	$\Sigma_{Bs,2,2}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[13]	$\Sigma_{clus}[13]$	Half-t( $\nu, A_{ph}$ )
Variance Bs[2,3]	$\Sigma_{Bs,2,3}$	Half-t( $\nu, A_{pm}$ )	Covariance Phot[14]	$\Sigma_{clus}[14]$	Half-t( $\nu, A_{ph}$ )
			Covariance Phot[15]	$\Sigma_{clus}[15]$	Half-t( $\nu, A_{ph}$ )

# Chapter 4

## Results

In this Chapter, I characterise the methodology detailed in Chapter 3 and then I applied it to the Pleiades DANCe DR2 data set (Sect. 2.7). To characterise the methodology as a classifier, I measure its precision and accuracy when applied on synthetic data where the true members of the cluster are known. With this characterisation, I am able to obtain an optimal probability threshold, but only for classification purposes. Afterwards, I apply the methodology to the Pleiades DANCe RDR2 and I find the candidate members of the cluster using the optimal probability threshold. Then, I compare these candidate members with those found by previous studies.

Later, I analyse the main results of this work, those that fulfil the objective: the statistical distributions that characterise of the cluster population. Then, I give the details of the spatial, velocity, luminosity and mass distributions. Finally, I end this Chapter describing the physical scenario of the evolution of the mass distribution of the Pleiades by comparing it with other mass distribution of younger and older clusters.

### 4.1 Performance of the classifier

As mentioned earlier, the main objective of the methodology of the BHM is the statistical characterisation of the NYC populations. However, as a by product, it also obtains the individual membership probability distributions of the objects comprising the data set. These membership probability distributions, together with a probability threshold, allow a direct classification of the objects into cluster and field members. The classification resulting from this procedure, as any other measured property, has an uncertainty. By evaluating this uncertainty under the results of synthetic data (in which the true members are known) we are able to measure the accuracy and precision of the classification process as a function of the probability threshold used. This section explains how an objective probability threshold can be found by maximising the accuracy of the classifier.

To measure the accuracy of our classifier, I test it over synthetic data sets that resemble

the real RDR2 (Section 2.7.2). An ideal test to our classifier will be to apply it over well known dataset in which tags of cluster and field members were already present. However, if we may have access to these tags, a classifier may not be needed. The Pleiades cluster being one of the most studied cluster in history, it is the NYC with most of these tags (see Section 4.2). This is one of the reasons for which we decided to benchmark our methodology on it. In spite of the large number of candidate members for the Pleiades clusters, the synthetic data and its true tags are still needed. The reasons are the following. First, the list of candidate members provided by the literature is not infallible. We can never be sure that this list is complete and unpolluted. Here it is important to note that the astrophysical domain of the phenomenon marks a very important distinction with common supervised classification methods. A perfect and real training set is not available. It must be created from simulations. Second, even if the most probable candidate members from the literature were used as a training set (as done for example by Sarro et al. 2014) the very faint end of the magnitude distribution (represented by brown-dwarfs) is still a *terra incognita* where candidate members are scarce or not even exist.

Thus, to overcome the problem of the true tags, we decided to create synthetic data sets. These synthetic true tags, and therefore the results obtained from them, rely on the assumption that our cluster and field models resemble the real data. I am aware that these models are far from perfect, but so far this assumption provides the best option. Although this assumption enable us to quantify the internal consistence (precision and accuracy) of our classifier, it does not give any indication about possible biases in the model. To explore this possibility, in the next Section, I compare our real data classification results with those of the literature.

The random nature of the synthetic data sets demands the repetition of the data sets and results. This repetitions avoid any bias caused by excursions of the random number generator (bad luck), and more importantly, they allow us to compute the uncertainty in the accuracy and precision of the classifier. As explained before, the methodology presented in this work is computing demanding. Thus, to be able to repeat at least five times the results of the synthetic data sets, we further reduce the data set size. The inference process in a data set of  $10^4$  objects demands almost week of computing time. Thus, we decided that such data set provided a good compromise between computing time and number of objects. Furthermore, to provide a better estimate of the contamination rate on the results over the real RDR2 data set (the one with the  $10^5$  objects) we choose to work with the  $10^4$  objects with high membership probability according to Bouy et al. (2015). Since this sample is relatively more entangled with the cluster than that of the RDR2, we assume that the contamination rate we measure on this sample will be comparable or even higher than that hypothetically obtained over the  $10^5$  RDR2.

Briefly, to create the synthetic data set, the procedure is the following. First, using the methodology of the previous Chapter, I obtain a sample of the posterior distribution of

the parameters in the model given the  $10^4$  real data set. Then, I choose the particle with highest posterior probability as the MAP estimate of the posterior distribution. Using this particle positions in the parametric space, I generate five synthetic data sets of  $10^4$  objects each. Then, I tag these objects according to their parent population: cluster or field. Afterwards, using the vector of synthetic values of each object, I assign their uncertainties and missing value entries (more details below).

Finally, I run the methodology over the five synthetic data sets, and obtain a sample of the individual membership probability distributions of each synthetic object. Then, I compare the true tags with the measured ones as function of the probability threshold.

To further test the performance of the classifier, I apply it on a synthetic data set in two different cases. In the first case the data set is one of the five synthetic ones. Thus it contains objects with missing value entries in their vector of observables. In the second case the data set is the same as in the previous case, but this time all the objects have fully observed vectors (i.e. do not have missing value entries). The comparison of the results rendered by these two cases allows us to quantify the impact of missing values.

Objects in the Pleiades DANCe DR2 have full observed proper motions vectors. Since missing values appear only in the photometric vectors, in synthetic data sets I only include missing values in the photometry. Only  $\sim 1\%$  of the objects in the RDR2 have full observed vectors (i.e. no missing value entries). Furthermore, the distribution of missing entries is not random and depends on the magnitudes of the objects (see Fig. 3.2). Therefore, to better reproduce this distribution in the synthetic data sets, for each synthetic datum, I use the mask of missing entries of one of its closer neighbours in the real data set. Here, distance is measured in the euclidean sense. If I were to use the missing value mask of the nearest neighbour in the real data set, then I will obtain a biased sample in which objects with full observed vectors (with no missing entries) will be underestimated. This is the inevitable consequence of the fact that euclidean distances measured in subspaces, which result from removing the dimensions of the missing entries, are small, or at most equal, than those measured in the complete space.

To reproduce the mask of missing entries, I choose from among those masks of the closer neighbours in the available CMDs:  $\{K_s, J - K_s\}$ ,  $\{J, J - H\}$ ,  $\{K_s, H - K_s\}$ ,  $\{J, Y - J\}$ ,  $\{K_s, i - K_s\}$ . These CMD are formed with the bands and colours of the magnitudes with fewer missing values, in decreasing order (see Table 2.2). The mask of missing entries to reproduce in the synthetic objects is chosen as follows. First, for each CMD subspace I find the set of objects, from the real data set, with fully observed entries, I call it  $C_{or,i}$  and its fraction from the total,  $f_r$ . Then, I take a random sample from the synthetic data whose fraction,  $f_s$  equals  $f_r$ . For objects in this sample I assign the missing value pattern of the nearest neighbour in  $C_{or,i}$ . I repeat this procedure for the rest of the CMDs. In this way, the synthetic data sets have fractions of objects with and without missing entries, similar to those of the objects in the real data set.

To assign uncertainties I proceed as follows. For uncertainties in the proper motions I use those of the nearest neighbour from the real data set. If I were to use the nearest neighbour scheme for uncertainties in the photometry, then I will obtain biased results. These uncertainties will be biased towards those of the less precise measurements. Again, this is a consequence of the missing entries in the observable vectors. The euclidean metric results in the preferential choosing of objects with missing values. These missing values occur mostly at the faint end, where uncertainties are larger. Therefore, these uncertainties will be biased towards these larger values. To avoid this issue, I fit 8th degree Chebyshev polynomials to the uncertainties as a function of the magnitudes. Then, I use these polynomials to establish the photometric uncertainties of the synthetic data sets.

Once the synthetic data sets were created I run the methodology on each of them and recover the membership probability distribution of each object. Then, I classify each object as cluster or field member. To classify an object as a cluster member, the mode of its cluster membership probability distribution must be higher than the probability threshold. Otherwise it is classified as a field object.

The classifier performance was measured by counting the true positives (TP, cluster members correctly classified), true negatives (TN, field members correctly classified), false positives (FP, field members classified as cluster members) and false negatives (FN, cluster members classified as field members) recoveries as a function of the probability threshold. With them I calculate the following quantities: the true positive rate (TPR), which is the ratio of true positives over the sum of true positives plus false negatives, the contamination rate (CR), which is the ratio of false positives over the sum of false positives plus true positives, the precision or positive predictive value (PPV), which is the ratio of true positives over the sum of true positives plus false positives, and, the accuracy (ACC), which is the ratio of the sum of true positives plus true negative over the sum of true and false positives and negatives.

These are,

$$\begin{aligned} TPR &= \frac{TP}{TP + FN} \\ CR &= \frac{FP}{FP + TP} \\ PPV &= \frac{TP}{TP + FP} \\ ACC &= \frac{TP + TN}{TN + FN + TP + FP}, \end{aligned}$$

which are defined at each probability threshold. I use the results of the five synthetic data sets to quantify the uncertainties of the previous quantities.

In Fig. 4.1, I show the mean and uncertainty of the TPR and CR measured on the five synthetic data sets. The uncertainty is represented by the maximum deviation from the mean. Also, this Figure shows the TPR and CR measured on a synthetic data set with

fully observed objects (i.e. objects with non missing entries). As this Figure shows, the missing values have a negative impact in our classification. They diminish the TPR and increase the CR. This negative impact is expected since the observables we are using are highly discriminant in the classification process (see Section 2.7.1). Since cluster and field are highly entangled in the  $10^4$  objects synthetic samples, when one of these observables is missing the classification is more uncertain or it could even be biased. Interestingly, the CR above probability threshold 0.8 is independent of the missing values and remains low ( $\lesssim 5\%$ ). In spite of the negative impact of missing values, the methodology delivers low contamination rates ( $\lesssim 5 - 10\%$ ) and high recovery rates ( $\lesssim 90 - 96\%$ ) for probability thresholds in the  $0.5 - 0.9$  range.

In Figure 4.1, I also show the CR and TPR of Sarro et al. (2014) (reported in their Table 4). Previous to discuss the differences between both works I inform the reader about the unfairness of this comparison. First, in the works of Sarro et al. (2014) and Bouy et al. (2015), the generative models are constructed using only fully observed objects (i.e. without missing entries), these objects represent only  $\sim 1\%$  of our RDR2. Afterwards, they apply those models to all the objects in the DANCe DR2 data set (i.e. objects with and without missing entries). Thus, their results are more similar to those I find on the synthetic data set with only fully observed objects (blue lines in Fig. 4.1). Second, the synthetic data sets in which both works measure the TPR and CR are different. They are constructed with different generative models, different number of objects, and different missing value distributions.

From the comparison between Sarro et al. (2014) TPR and CR, with the ones I find on the synthetic data set comprising only fully observed objects, we see the following. The TPR of both works agree within the uncertainties. The CR of both works agree, within the reported uncertainties, for probability thresholds below 0.8. At higher probability thresholds, our methodology delivers lower CR. Even in the case of the unfair comparison between the TPR and CR of Sarro et al. (2014) with those I measure on the synthetic data sets including objects with missing entries, our CR outperforms that reported by Sarro et al. (2014) at the small price of a lower ( $\sim 4\%$ ) TPR.

Now, I describe the procedure to set an optimal probability threshold. This probability threshold, although not needed to obtain the posterior distribution of the parameters modelling the cluster population, is needed, however, to objectively classify an object as a cluster member. I establish this threshold using only the synthetic data sets containing objects with missing entries in their vector of measurements. The approach I use to set this probability threshold is that of the maximum accuracy (ACC) for the classification.

Figure 4.2 shows the ACC and the PPV of the classifier when it is applied on synthetic data sets containing objects with missing value entries. The lines and the grey regions depict, respectively, the mean and the maximum deviations of the results of the five synthetic data sets. The maximum deviation is used as a proxy for the uncertainty.

The highest mean accuracy,  $\text{ACC}=96.5 \pm 0.1\%$ , happens at probability threshold  $p_t = 0.84$ . Thus, I choose it as the optimal classification threshold. At this value, the CR is  $4.3 \pm 0.2\%$ , the TPR is  $90.0 \pm 0.05\%$ , and the PPV is  $95.6 \pm 0.2\%$ .

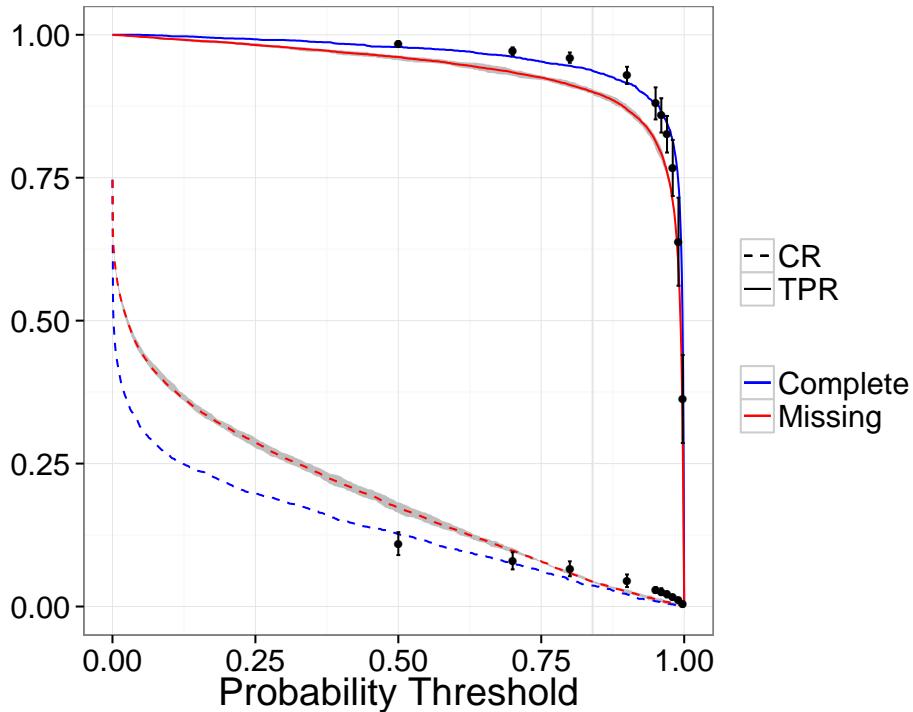


Figure 4.1: The mean TPR (solid line) and CR (dashed line) resulting from five synthetic data sets including objects with missing entries (red lines). Also the TPR and CR resulting from a synthetic data set comprising only objects with fully observed vectors (blue lines). The shaded regions (grey) show the uncertainties computed from the five synthetic data sets. The black dots show the TPR and CR reported by [Sarro et al. \(2014\)](#) for their model. See text for warnings on this comparison. Reproduced from Figure 3 of [Olivares et al. \(2017\)](#), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, Vol. submitted.

We further investigate the impact that objects with missing value entries in their observations have on our methodology. In specific, I analyse possible biases introduced by these objects. To do this, I compare the membership probabilities, summarised by the mode, recovered after inferring the model on two synthetic data sets of. These two data sets are identical except that in one of them some entries in the vector of observables where masked as missing (using the procedure previously described).

In Fig. 4.3, I compare the mode of the membership probabilities. The horizontal axis shows the membership probabilities of the data set with fully observed objects (I call this

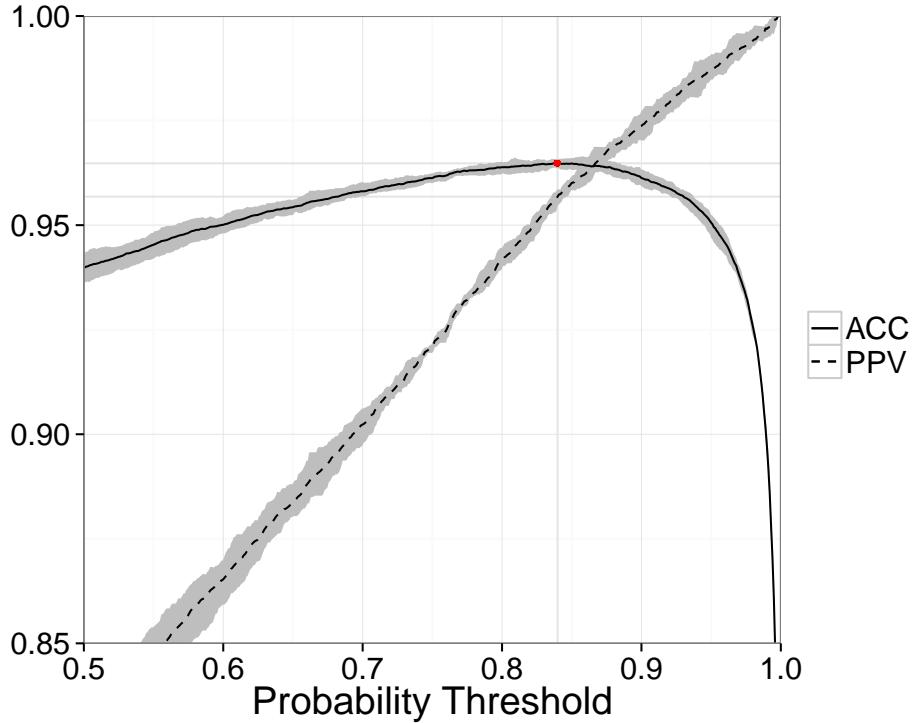


Figure 4.2: Mean accuracy (ACC, solid line) and precision (PPV, dashed line) of the classifier as a function of probability threshold. The shaded regions shows the uncertainties computed from the five synthetic data sets. The higher accuracy is obtained at  $p_t = 0.84$  (red dot). Reproduced from Figure 3 of [Olivares et al. \(2017\)](#), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, Vol. submitted.

case the complete one). The vertical axis shows the membership probabilities of the same objects but in which some entries were masked as missing (I call this case the Incomplete one). As can be seen in this figure, the missing values impact our results by spreading the membership probabilities. Ideally, we would like to recover membership probabilities following the line of slope one. This is the case of some fully observed objects (red squares) in the data set containing objects with missing entries. The most striking deviations come from those objects with the  $CI$  masked as missing (enclosed in black). The BHM methodology uses the *true CI* to prescribe the *true* photometry. Also, it uses the observed  $CI$  to constrain the marginalisation integral of the *true CI*. Thus, as expected, a missing  $CI$  produces a spread in the membership probability.

The objects with a missing  $CI$  show two different behaviours. In one case, there are objects with membership probabilities from the fully observed data set (horizontal axis) that have an overestimated membership probability in the missing entries data set (vertical

axis). In the other case, there are objects that have underestimated probabilities in the missing entries data set. These objects correspond to those seen in the combed area below the line of unit slope.

Objects in the first case increase the CR, and their effect can be seen by the difference between red and blue dashed lines in Fig. 4.1. On the other hand, the objects in the second case diminish the TPR, their effect can also be seen by the difference between red and blue solid lines in Fig. 4.1. The increase in CR reaches its maximum near probability zero in the horizontal axis (Complete case) and goes to zero at probability thresholds of  $\sim 0.9$ . Therefore, the impact this increased CR has in our results is marginal. For example, at the optimal probability threshold  $p_t = 0.84$ , the increase of CR due to objects with missing entries represent only 1.8%. This correspond to the objects in box region of Fig. 4.3 (upper left corner). However, the objects in the second case, those that diminish the TPR, represent the typical unavoidable loss of members due to their missing entries. These amount to a 4% loss in the TPR, at the optimal probability threshold,  $p_t = 0.84$ .

The bias introduced in the recovered membership probabilities due to objects with missing value entries, can be quantified using the root-mean-square (rms) of the difference between the means of the two recovered membership probabilities (Complete and Incomplete cases). The total rms is 0.12. On the one hand, fully observed objects in both data sets (Complete and Incomplete cases) have a rms of only 0.02. On the other hand, objects with missing entries, excluding those with missing  $CI$ , have a rms of 0.08. The rms of objects lacking the  $CI$  is 0.14. The previous effects show an overall agreement between results on data sets with and without objects with missing entries. Nonetheless, care must be taken when dealing with individual membership probabilities. An object with a missing value in the  $Y, J, H$  and  $K_s$  may have a diminished membership probability (with a rms of 0.08), while an object with a missing  $CI$  may show an increased membership probability (with a rms of 0.14).

However, as have been mentioned before, the methodology described in this work aims at the statistical distributions of the cluster population. The individual membership probabilities are just a useful by product. The methodology develop here, though, works by ensuring that each object contributes to the posterior distribution of the parameters modelling the cluster population, proportionally to its cluster membership probability. In this sense our results are free of any possible bias introduced by cuts in the membership probability. Nevertheless, there is still contamination. In particular, that arising from objects with missing entries. This contamination in the statistical distributions that we aim to obtain must be quantified. To do this, I compute the expected value of the CR found in this section. It is  $\langle CR \rangle = 5.8 \pm 0.2\%$ . In this expected value, each CR contributes proportionally to the probability threshold at which it is measured. Since the vast contribution to this CR comes from probability thresholds below 0.2 (see Fig. 4.1), the expected value of the CR remains low.

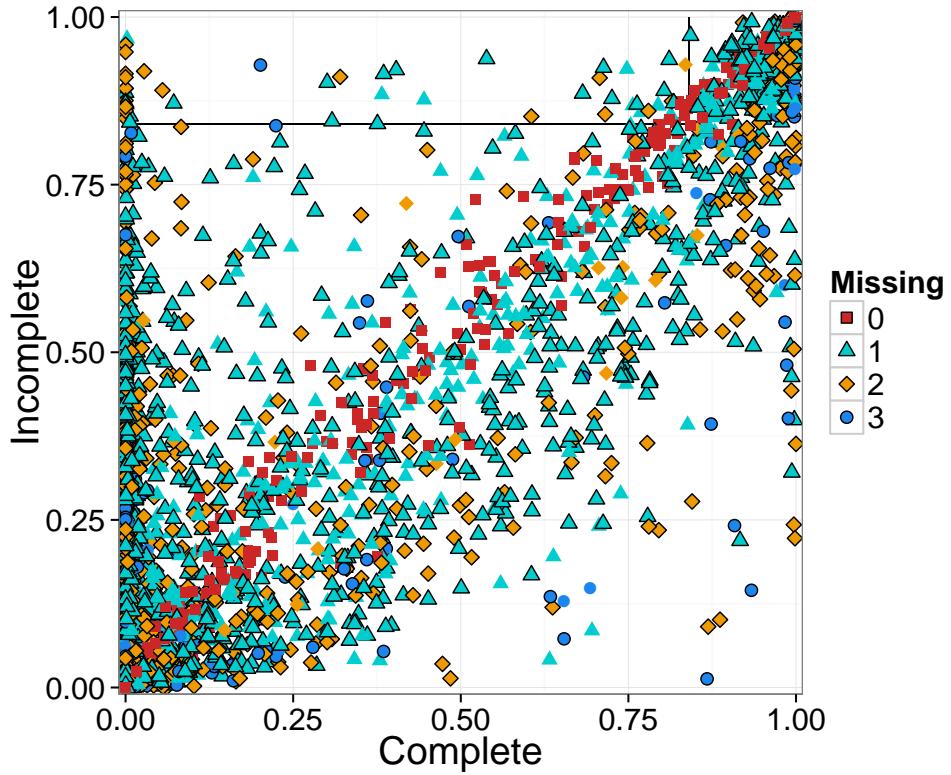


Figure 4.3: Comparison between the cluster membership probabilities recovered from the synthetic data set with objects having missing value entries (vertical axis, labeled Incomplete), and, the synthetic data set with fully observed objects (horizontal axis, labeled Complete). The colour and shape indicate the amount of missing entries. The symbols enclosed in black indicate a missing  $CI$ . The top left box contains objects considered as contaminants due to missing values at the probability threshold  $p_t = 0.84$ . Reproduced from Figure 4 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, Vol. submitted.

In statistical science, in machine learning particularly, is sometimes useful to analyse the performance of a binary classifier by the receiver operating characteristic curve, the ROC curve. It plots a visual diagnostic of the ability of a classifier to perform its job. The ROC curve plots the TPR as a function of the FPR for all possible values of the probability threshold. A perfect classifier would be that in which the TPR=1 and the FPR=0 for some probability threshold. On the other hand, a random classifier would be that with TPR=FPR at all probability thresholds. Such classifier has a line of slope one in its ROC curve. Furthermore, the quantitative diagnostic for a binary classifier is the area under the ROC curve (AUC). As its name indicate, the AUC is the integral of the ROC curve. Thus a random classifier has a AUC of one half, while a perfect classifier has a AUC=1. In Fig. 4.4, I show the ROC curve for our classifier when applied over synthetic data containing objects with missing entries. It is the ROC of one of the five synthetic realisations described throughout this section. As can be seen from this Figure our classifier does an excellent job, with an AUC=0.992.

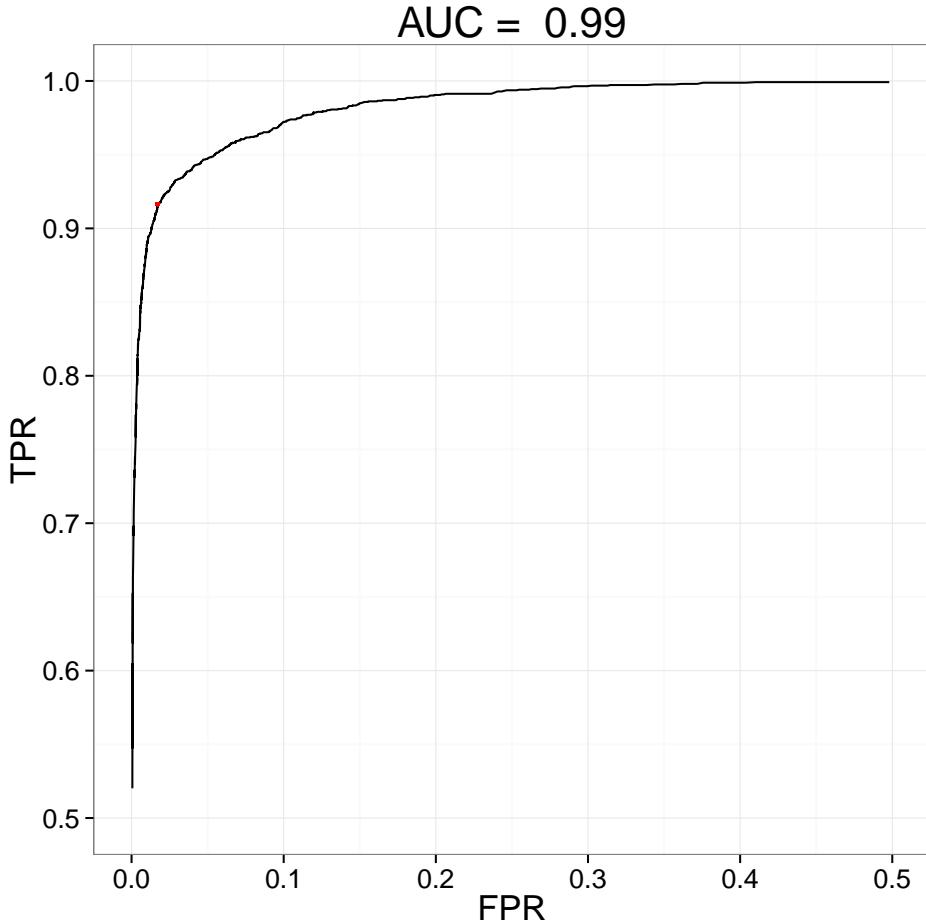


Figure 4.4: ROC curve of the BHM by-product classifier when applied on the synthetic data set containing objects with missing entries. As can be seen, the AUC=0.992 diagnose it as an excellent classifier.

## 4.2 Comparison with the literature

The BHM methodology, described in Chapter 3, is applied to the Pleiades DANCe RDR2. Then, I compare the recovered membership probabilities with those reported in the literature. The first comparison I present is that of our candidate members with those in the two lists of candidate members of Stauffer et al. (2007). Then, I proceed to compare our membership probabilities with those found by Bouy et al. (2015). Finally, I compare the candidate members recovered by the BHM with those found by Rebull et al. (2016). These authors obtain their list of candidates based on photometric variability. This observable is not present in our set of observables, for this reason this comparison represent an extra source of external validation.

### 4.2.1 Candidate members from Stauffer et al. (2007)

Stauffer et al. (2007) published two list of candidate members. The first one contains 1417 objects compiled from the literature (see Table 2 of the mentioned work). These objects were classified as candidate members by several authors. As Stauffer et al. (2007) mention, this list is inhomogeneous, incomplete and certainly includes non-members. I refer to this list as ST1. Their second list contains 55 candidate members (see Table 5 of the mentioned work). Stauffer et al. (2007) found these members using infrared photometry and proper motions. I refer to this list as ST2.

Cross matching (at CDS<sup>1</sup>, within 1 arcsec radius) the previous two lists with the Pleiades DANCe DR2 catalogue (Bouy et al. 2015), I find that only 1384 and 54 of the ST1 and ST2 lists have a counter part in the Pleiades DANCe DR2 catalogue, respectively.

Concerning our list of candidate members, after cross matching it with the two lists of Stauffer et al. (2007), ST1 and ST2, I recover 1146 and 34 of the candidate members, respectively. Compared to the candidate members of Bouy et al. (2015), our BHM recovers 28 more candidate members in ST1 and the same in ST2.

As mentioned before, the ST1 list is an exhaustive compilation of Pleiades members. It contains objects that were classified, at some point in history, as Pleiades candidate members, even when their membership probability are as low as 0.1 (Stauffer et al. 2007). For this reason I will not analyse the details of 238 rejected objects of ST1. It suffices to show that these rejected objects lie far from the cluster photometric or proper motions loci, as shown in Fig. 4.5.

On the other hand, from the 20 objects in ST2 that are rejected by the BHM, 18 of them lie below the cluster photometric sequence and far from the proper motion locus, see Fig. 4.6. The remaining two (DANCe IDs: J034552.57+235145.9 and J034543.47+233851.5), although have observed photometric vectors compatible with the clusters sequence, their proper motions still are far from the cluster centre (with  $\mu_\alpha \cos(\delta) \sim 40 \text{ mas} \cdot \text{yr}^{-1}$ ), see

---

<sup>1</sup> Using the service <http://cdsxmatch.u-strasbg.fr/xmatch>

Fig. 4.6.

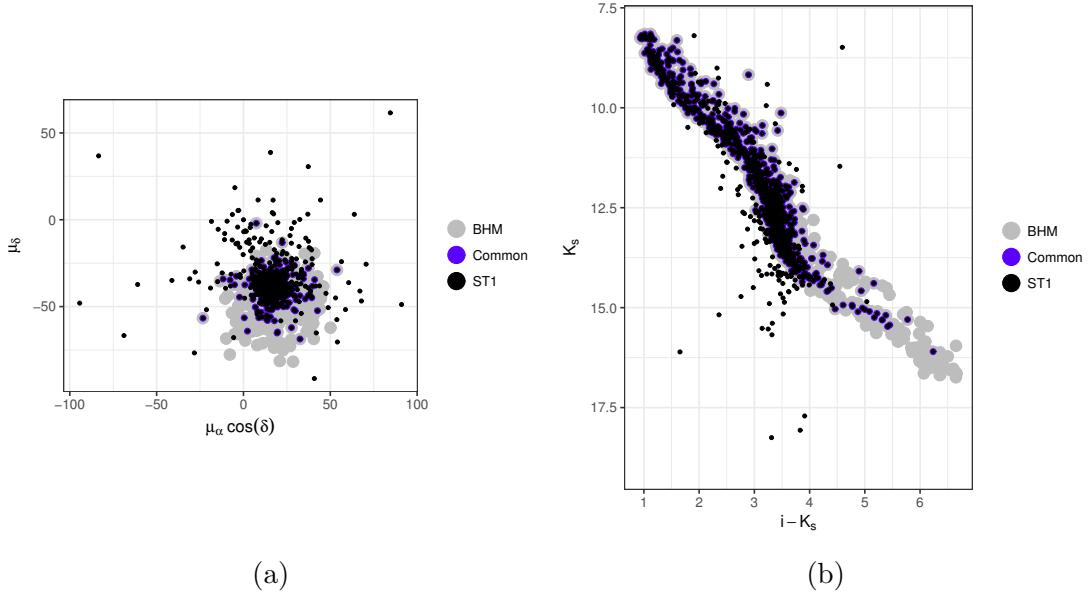


Figure 4.5: Proper motions (a) and  $K$  vs  $i - K$  CMD (b) of the ST1 candidate members in the DANCe DR2 catalogue (black). Also shown, the objects classified as candidate members in the BHM (grey), and in both ST1 and BHM (blue).

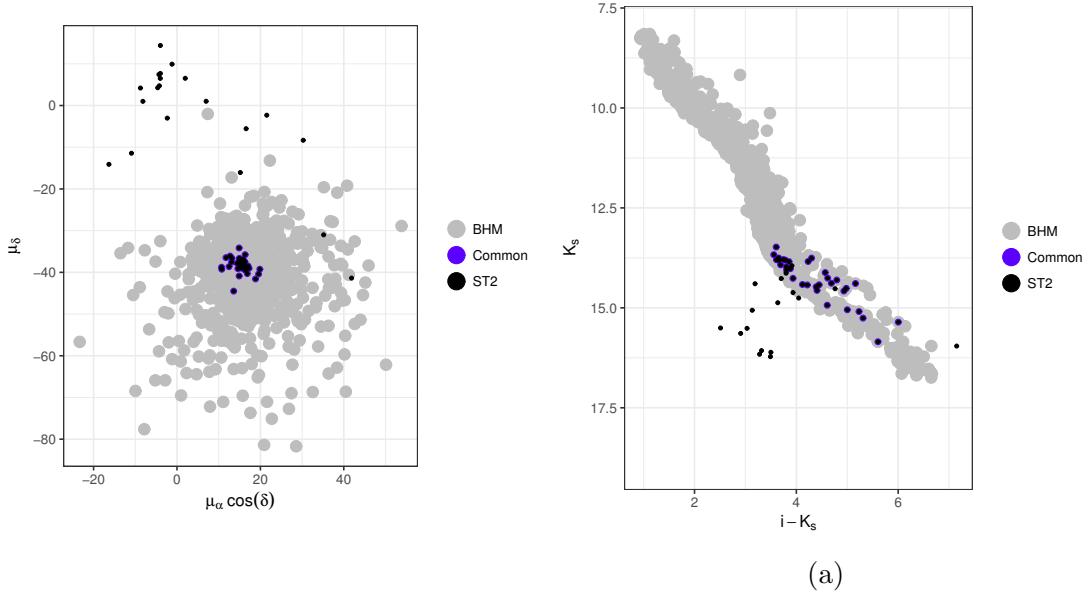


Figure 4.6: Proper motions (a) and  $K$  vs  $i - K$  CMD (b) of the ST2 candidate members in the DANCe DR2 catalogue (black). Also shown, the objects classified as candidate members in the BHM (grey), and in both ST2 and BHM (blue).

#### 4.2.2 Candidate members from Bouy et al. (2015)

The fact that the work of [Bouy et al. \(2015\)](#) and the present one use the same DANCe DR2 data set (although our model is constructed on the RDR2), allow me to directly

compare the membership probabilities of both works. Furthermore, this comparison can be extended to all objects in the data set and not just to the candidate members of the Pleiades cluster. Since [Bouy et al. \(2015\)](#) reported only a statistic of the membership probability distributions, to do a fair comparison, I summarise the membership probability distributions recovered by the BHM with the mode.

Using the optimal probability threshold of 0.84 to classify the cluster members, we can see that, as shown by Fig. 4.7, both methodologies agree on the outstanding 99.6% of the classified objects. Concerning just the candidate members, the agreement is still high,  $\sim 90\%$ . In the following I discuss the 10% discrepancies.

The candidate members of [Bouy et al. \(2015\)](#) that the BHM rejects, which I call the rejected ones, are shown in lower right box of Fig. 4.7. They amount to 12% and 12.5% of total number of candidate members recovered by [Bouy et al. \(2015\)](#) and the present work, respectively. The 12% figure is 4.7% higher than the contamination rate (CR) reported by [Sarro et al. \(2014\)](#) ( $7.3 \pm 1.4\%$ ). It indicates that some true cluster members must be within the rejected objects. Also, the 12.5% figure is 2.5% higher than the 10% loss rate of the BHM (TPR=90%, see Section 4.1), which indicates that some contaminants must be within the rejected objects.

Now, I analyse these objects with further detail. As it is shown in Figs. 4.8 and 4.9, the rejected objects have proper motions uncertainties with median  $\tilde{\mu}_\alpha, \tilde{\mu}_\delta = \{3.19, 3.20\} \text{ mas} \cdot \text{yr}^{-1}$ . This value is more than four times larger than that of the common candidate members (those objects classified as members by both works, see top right corner of Fig. 4.7), which have median  $\tilde{\mu}_\alpha, \tilde{\mu}_\delta = \{0.68, 0.68\} \text{ mas} \cdot \text{yr}^{-1}$ . Among the rejected ones, those with a relatively high membership probability occur mostly at the middle of the cluster photometric sequence (green squares of Fig. 4.9). On the other hand, those with lower membership probabilities occur at the bright and faint ends (blue and red triangles of Fig. 4.9, respectively). Furthermore, the proper motions uncertainties of the rejected objects at the bright, middle and faint ends of the cluster photometric sequence, have medians of  $\tilde{\mu}_\alpha, \tilde{\mu}_\delta = \{4.3, 4.2\} \text{ mas} \cdot \text{yr}^{-1}$ ,  $\tilde{\mu}_\alpha, \tilde{\mu}_\delta = \{2.4, 2.4\} \text{ mas} \cdot \text{yr}^{-1}$  and  $\tilde{\mu}_\alpha, \tilde{\mu}_\delta = \{3.4, 3.4\} \text{ mas} \cdot \text{yr}^{-1}$ , respectively. These figures are approximately 6, 4 and 5 times larger, respectively, than those of the candidates in common. These large uncertainties produce a proportional spread of the cluster likelihood. This could reduce the membership probability of these objects.

Furthermore, the bright and faint regions of the cluster photometric sequence coincide with those where objects with missing entries are more frequent (see Fig. 3.2). I stress the fact that the model of [Bouy et al. \(2015\)](#) was constructed using only objects with fully observed vectors of measurements, which represent less than 1% of the number of objects used by the BHM. Using only fully observed objects, underestimates the field density in the regions where the missing values are more frequent. Underestimating the field likelihood increases the cluster to field likelihood ratio. Therefore it increases also

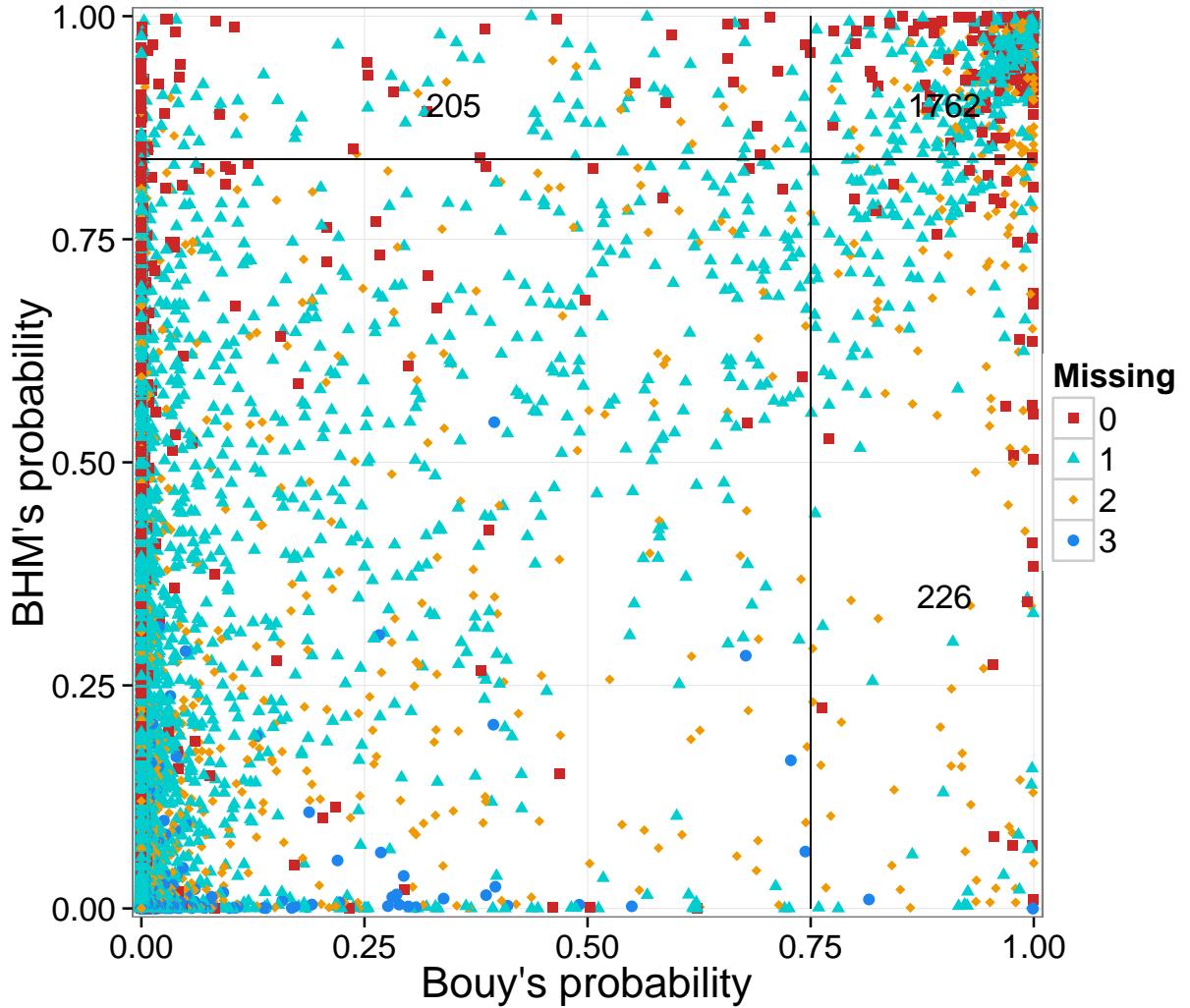


Figure 4.7: Mode of the membership probabilities recovered by the BHM compared to those of Bouy et al. (2015). The lines show the 0.75 and  $p_t = 0.84$  probability thresholds used in both works. The numbers indicate our new candidate members (top left), the ones we rejected (bottom right), and the common ones (top right). Reproduced from Figure 11 of Olivares et al. (2017), *The Seven sisters DANCe III: Hierarchical Bayesian model*, A&A, Vol. submitted.

the cluster membership probabilities. This may explain the relatively high membership probabilities reported by Bouy et al. (2015) compared to those of the BHM. In particular, those in the faint region of the photometric cluster sequence.

Therefore, I consider that both previous phenomena, the large proper motion uncertainties of the rejected objects together with the underestimated field likelihood in Bouy et al. (2015) model, are responsible for the relatively lower BHM membership probabilities of the rejected objects. However, these objects can not be discarded as potential true

cluster members. To discard this possibility lower proper motion uncertainties and fewer missing values are needed. Future steps will be taken to try to solve this issue.

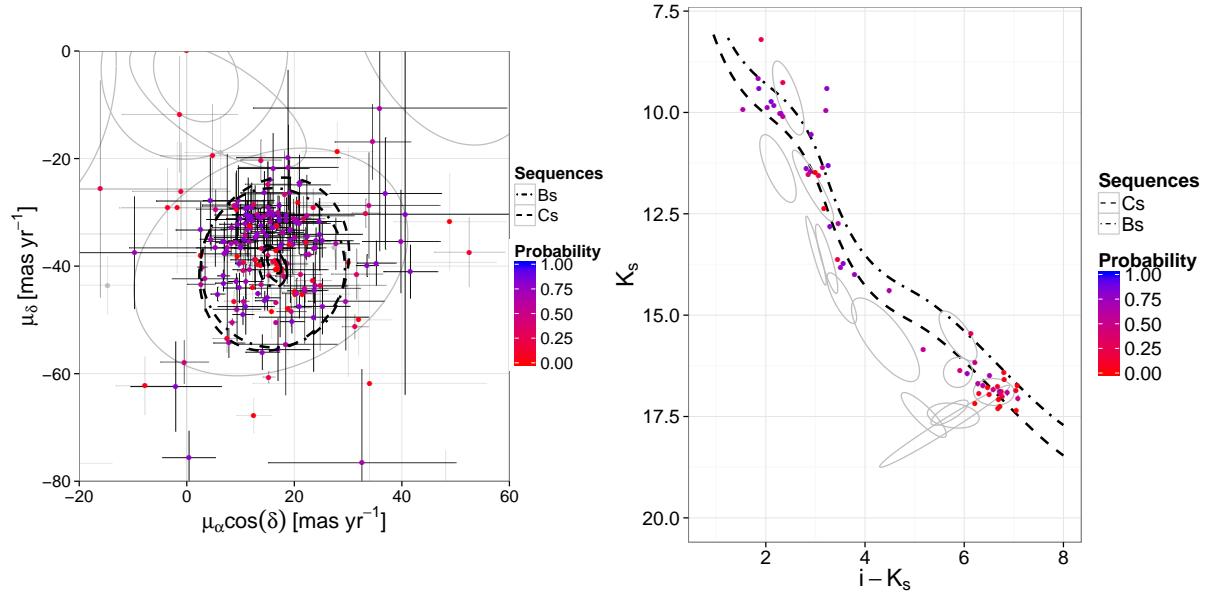


Figure 4.8: Proper motion (left) and  $K_s$  vs.  $i - K_s$  CMD (right) showing the candidate members of Bouy et al. (2015) rejected by the BHM. Reproduced from Figure 13 of Olivares et al. (2017), *The Seven sisters DANCe III: Hierarchical Bayesian model*, A&A, Vol. submitted.

On the other hand, the new candidate members found by the BHM, shown in the upper left box of Fig. 4.7, amount to 10% of the Bouy et al. (2015) candidate members. This figure is higher than the  $\sim 3.5\%$  of missing rate (1-TPR) reported by Sarro et al. (2014). Also, these new candidates amount to 10% of the BHM recovered candidate members. This value is larger than the 4.3% CR reported in Section 4.1. These two larger figures may indicate that some truly new discoveries may be within these lists of new candidate members. In Figs. 4.10 I show the proper motions and  $K_s$  vs  $i - K_s$  CMD of these new candidate members.

The new candidate members have proper motions uncertainties whose median,  $\tilde{\mu}_\alpha, \tilde{\mu}_\delta = \{1.41, 1.41\}$  mas  $\cdot$  yr $^{-1}$ , is two times larger than those of the candidate members in common with Bouy et al. (2015). Also, as shown by Fig. 4.10, the majority of the new candidate members, 166, have probabilities lower than 0.95, are located in a halo around the locus of the cluster proper motions, and on top of the cluster photometric sequence in the  $K_s$  vs  $i - K_s$  CMD. On the contrary, the new candidate members with probabilities higher than 0.95, which are 39, lie in the centre of the cluster proper motions and fall above the cluster sequence in the  $K_s$  vs  $i - K_s$  CMD. Thus, I hypothesise that: i) Objects whose photometry is compatible with the cluster sequence but are in the proper motions halo, have higher membership probabilities in our methodology due to the increased flexibility of the cluster proper motions model: it now has four gaussians instead of the two of Bouy et al. (2015).

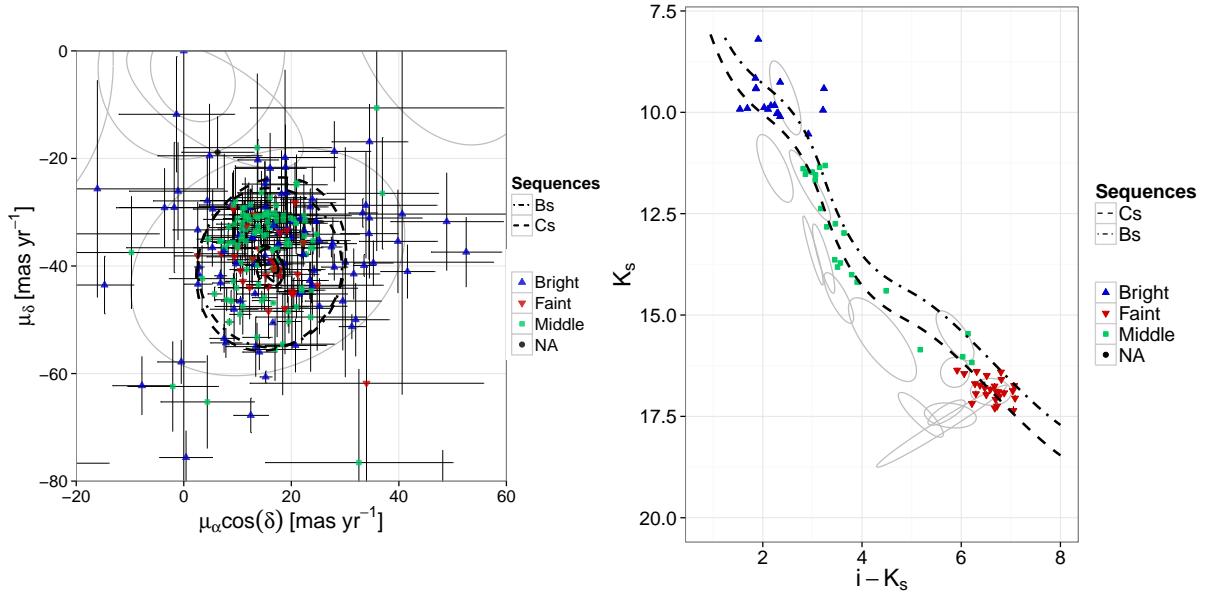


Figure 4.9: Proper motion (left) and  $K_s$  vs.  $i - K_s$  CMD (right) showing the candidate members of [Bouy et al. \(2015\)](#) rejected by the BHM. The colours and shapes are a proxy for their  $K_s$  magnitude. Reproduced from Figure 14 of [Olivares et al. \(2017\)](#), *The Seven sisters DANCe III: Hierarchical Bayesian model*, A&A, Vol. submitted.

And ii) objects near the centre of the cluster proper motions but located above the cluster photometric sequence sequence, are multiple systems (probably triple systems which can amount to 4% of the population [Duquennoy & Mayor 1991](#)) with an increased membership probability due to our more flexible photometric model of the cluster and equal-mass binaries sequences.

Summarising, the discrepancies between the BHM membership probabilities and those reported by [Bouy et al. \(2015\)](#) arise from subtle but important differences. The first difference is the treatment of object with missing entries. Although [Bouy et al. \(2015\)](#) report membership probabilities for this kind of objects, their field and cluster models were constructed discarding them. This may have biased their results. Taking into account objects with missing entries has two consequences. First, the photometric model is more accurate than any other model that discards objects with missing entries. This effect is important in the regions where these objects are more frequent. Second, the use of objects with missing entries in the construction of the cluster model allow us to include the information of good candidate members that were otherwise discarded a priori. A second difference with the model of [Bouy et al. \(2015\)](#), is the higher flexibility of our cluster model. It allows us to increase the membership probability of the previously discarded candidates.

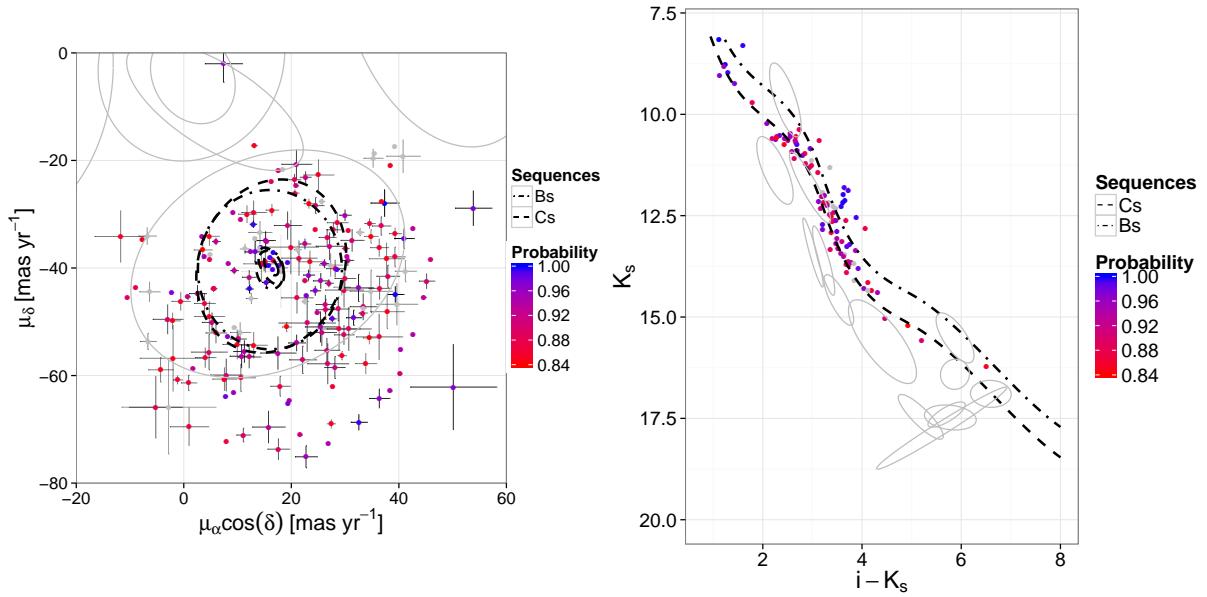


Figure 4.10: Proper motion (left) and  $K_s$  vs.  $i - K_s$  CMD (right) showing the new candidate members found in this work. Reproduced from Figure 12 of Olivares et al. (2017), *The Seven sisters DANCe III: Hierarchical Bayesian model*, A&A, Vol. submitted.

### 4.2.3 Candidate members from Rebull et al. (2016)

After cross matching (at CDS with a 0.5 arcsec radius) the list of candidate members from Rebull et al. (2016) (their Table 2, here after RT1) with the DANCE DR2, I find that 758 out of the 759 objects have a counter part in the DANCE DR2 catalogue. The 91% of these objects (690 of them) are candidate members in the BHM. Under the assumption that the candidate members in RT1 are indeed true members, which may not be true, the ratio of recovered members is even better than the  $TPR = 90 \pm 0.2\%$  reported in Section 4.1. These objects are shown in Fig. 4.11.

On the other hand, after cross matching (at CDS with a 0.5 arcsec radius) the list of 154 objects that Rebull et al. (2016) classify as non-members (their Table 6, here after RT2) with the DANCE DR2, I find that all these objects have a counter part on the DANCE DR2. The 21% of objects in the RT2 list (33 of them) were classified as candidate members in the BHM. This is a value five times larger than the CR reported in Section 4.1 (CR=4.3%). However, we can not assume that the RT2 list comprises only non-members. First, these objects were at some point classified as members by other authors (Appendix B of Rebull et al. 2016). Second, not all of these objects have periods (only 20% according to Rebull et al. 2016). From the 33 objects classified as candidate members by the BHM, only nine of them have periods. It means that for the remaining 24 candidate members, these authors used other criteria to discard them as members. About their classification process, Rebull et al. (2016) say “This process was qualitative in the sense that we weighted all of the information in a subjective manner. However, the process was also extensive, with

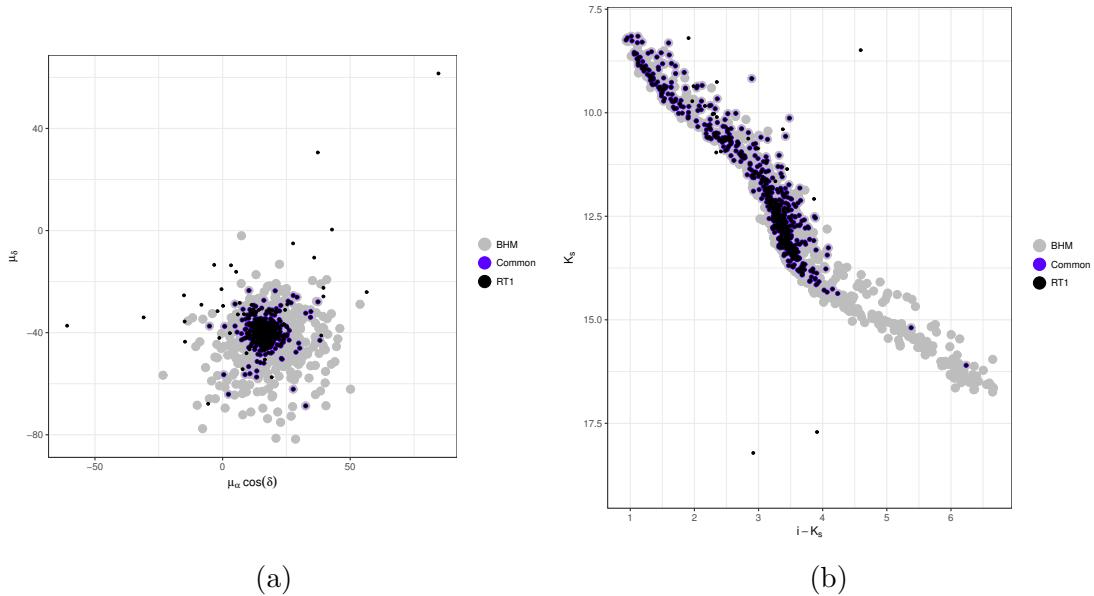


Figure 4.11: Proper motions (a) and  $K$  vs  $i - K$  CMD (b) of the RT1 candidate members in the DANCe DR2 catalogue (black). Also shown, the objects classified as candidate members in the BHM (grey), and in both RT1 and BHM (blue).

each star considered individually and with all available information considered in detail. [...] we believe that in the great majority of cases we have made the right decision.” In any case, the high rate of BHM candidate members found in this RT2 list may indicate that either our contamination rate is underestimated or that the criteria used by [Rebull et al. \(2016\)](#) are too restrictive. As can be seen in Fig. 4.12, these 33 objects have proper motions and photometric measurements consistent with those of the cluster. In order to clarify the status of these objects, further information is still needed.

### 4.3 The statistical distributions of the Pleiades cluster.

Now, I present the results of the statistical distributions that describe the cluster population, which are the main objective of the present work. These distributions result directly or indirectly from the posterior distribution of the parameters in our model. Indirectly means that I use them as parameters of other function (e.g. the mass distribution). Since we have 85 parameters in the BHM, I only discuss the posterior distributions of some of these parameters. In particular, those related with the velocity, luminosity and mass distributions. Nevertheless, in Table 4.1, I summarise the posterior distribution of the parameters in our model using the mode, also I use the 16th and 84th percentiles as a proxy for the uncertainty. The parameter names in this Table correspond to those given in Section 3.5 (at Table 3.1).

Also, for the sake of completeness, in Fig. 4.13 I depict the values the correlation

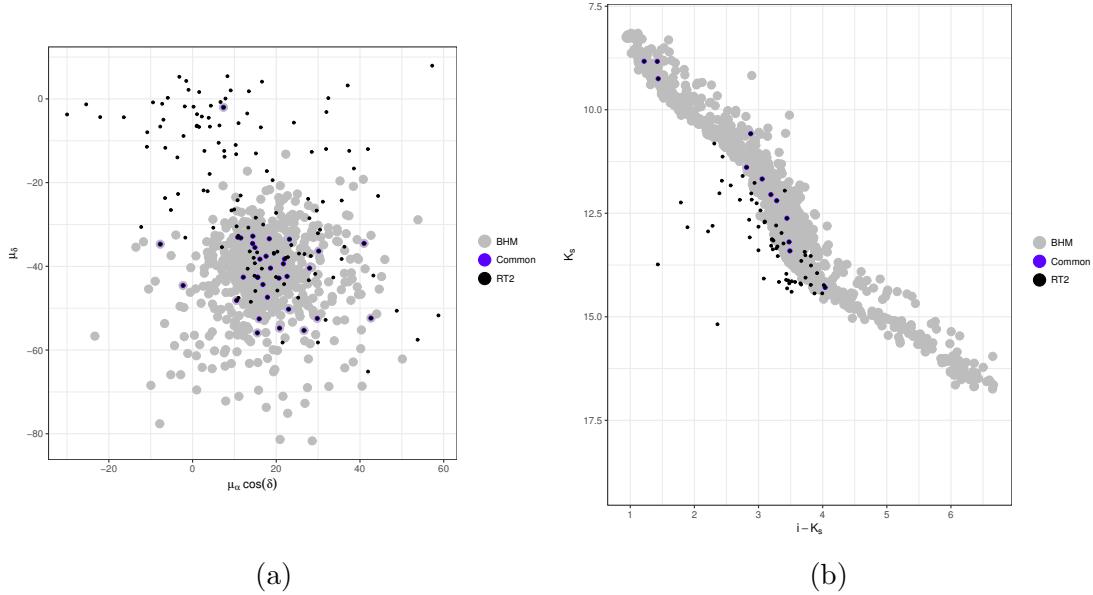


Figure 4.12: Proper motions (a) and  $K$  vs  $i - K$  CMD (b) of the RT2 in the DANCe DR2 catalogue (black). Also shown, the objects classified as candidate members in the BHM (grey), and those classified as non-members in the RT2 and as candidate members in the BHM (blue).

coefficients among the 85 parameters in the BHM. As can be seen from this Figure, the larger correlations appear among parameters describing the true  $CI$  distribution, and among these and almost the rest of the parameters. This is expected since the true  $CI$  is key parameter in the BHM. It is also interesting to notice that there is a strong correlation among the coefficients of the splines series modelling different magnitudes. For example, there is a strong correlation among the fourth coefficients of the splines. These correlations are expected since the shape of the cluster sequence is similar in the four CMD.

Table 4.1: Mode, 16 and 84 percentiles of each parameter posterior distribution.

Parameter	Mode	$p_{16\%}$	$p_{84\%}$
Field fraction	0.967958	0.967474	0.969462
Cs fraction	0.909371	0.895137	0.919434
Cs PM fraction 1	0.001347	0.001347	0.001347
Cs PM fraction 2	0.525735	0.508956	0.543462
Cs PM fraction 3	0.190321	0.159262	0.224475
Bs PM fraction 1	0.104021	0.103894	0.104108
Color fraction 1	0.085997	0.075534	0.094833
Color fraction 2	0.353973	0.193003	0.469717
Color fraction 3	0.277648	0.224469	0.289170
Color fraction 4	0.251190	0.105385	0.376304
Mean color 1	1.307689	1.259029	1.338972
Mean color 2	3.204578	2.841093	3.241497

Table 4.1: continued.

Parameter	Mode	$p_{16\%}$	$p_{84\%}$
Mean color 3	3.341761	3.321369	3.359906
Mean color 4	3.575646	3.518013	3.641340
Mean color 5	5.446573	5.366774	5.560569
Variance color 1	0.086389	0.076258	0.125984
Variance color 2	0.369646	0.275289	0.435338
Variance color 3	0.026526	0.026468	0.026565
Variance color 4	0.269562	0.269046	0.270325
Variance color 5	0.375299	0.322775	0.537458
Mean PM Cs[1,1]	16.305803	16.220500	16.384057
Mean PM Cs[1,2]	-39.683588	-39.767616	-39.469303
Variance Cs[1,1]	0.000000	0.000000	0.000000
Variance Cs[1,2]	0.000000	0.000000	0.000000
Variance Cs[1,3]	0.000000	0.000000	0.000000
Variance Cs[2,1]	193.231940	193.035542	193.511106
Variance Cs[2,2]	17.895759	6.705775	30.727887
Variance Cs[2,3]	259.457517	259.296878	259.651477
Variance Cs[3,1]	7.049380	6.007706	9.154742
Variance Cs[3,2]	-3.350959	-3.924016	-1.704487
Variance Cs[3,3]	11.685825	11.684455	11.687112
Variance Cs[4,1]	1.685699	1.608666	1.740720
Variance Cs[4,2]	-0.830553	-0.833982	-0.823643
Variance Cs[4,3]	3.590195	3.511369	3.833215
Mean PM Bs[1,1]	15.719802	15.521532	16.230056
Mean PM Bs[1,2]	-40.314746	-40.365824	-40.182113
Variance Bs[1,1]	339.480156	278.765965	392.206780
Variance Bs[1,2]	-0.421560	-0.443185	-0.404809
Variance Bs[1,3]	150.552074	56.130431	330.515498
Variance Bs[2,1]	6.512542	6.498438	6.521515
Variance Bs[2,2]	-1.421797	-2.214778	0.327916
Variance Bs[2,3]	10.696936	10.234819	11.259143
Coefficient [1,1]	6.862152	6.698804	6.962798
Coefficient [1,2]	12.549203	12.543435	12.559731
Coefficient [1,3]	10.638663	10.624438	10.646821
Coefficient [1,4]	16.260306	16.253606	16.280723
Coefficient [1,5]	16.864151	16.796169	16.907984
Coefficient [1,6]	21.087408	21.010410	21.183274
Coefficient [1,7]	23.279645	23.273198	23.290056

Table 4.1: continued.

Parameter	Mode	$p_{16\%}$	$p_{84\%}$
Coefficient [2,1]	7.622220	7.605694	7.625557
Coefficient [2,2]	11.527325	11.517715	11.567059
Coefficient [2,3]	10.211935	10.211547	10.212698
Coefficient [2,4]	15.602847	15.599241	15.621497
Coefficient [2,5]	16.098859	16.055647	16.153988
Coefficient [2,6]	19.338719	19.289969	19.399707
Coefficient [2,7]	21.813285	21.715478	21.957425
Coefficient [3,1]	7.574385	7.564066	7.590012
Coefficient [3,2]	10.970652	10.955195	11.009520
Coefficient [3,3]	9.485289	9.480620	9.494432
Coefficient [3,4]	15.111703	15.100806	15.126351
Coefficient [3,5]	15.343240	15.302140	15.391826
Coefficient [3,6]	18.632106	18.591372	18.649827
Coefficient [3,7]	19.654121	19.622414	19.678001
Coefficient [4,1]	7.529844	7.521514	7.539968
Coefficient [4,2]	10.860704	10.845059	10.895944
Coefficient [4,3]	9.334359	9.331793	9.337694
Coefficient [4,4]	14.722486	14.717526	14.741820
Coefficient [4,5]	14.954292	14.909027	15.015568
Coefficient [4,6]	17.636682	17.582166	17.705691
Coefficient [4,7]	18.461156	18.254644	18.788302
Covariance Phot[1]	0.128241	0.128140	0.128294
Covariance Phot[2]	0.028797	0.027836	0.032340
Covariance Phot[3]	0.007540	0.007387	0.007617
Covariance Phot[4]	0.001356	0.000907	0.001813
Covariance Phot[5]	-0.012226	-0.012248	-0.012174
Covariance Phot[6]	0.000000	0.000000	0.000000
Covariance Phot[7]	-0.009281	-0.009301	-0.009267
Covariance Phot[8]	-0.027990	-0.028689	-0.027546
Covariance Phot[9]	-0.029916	-0.030951	-0.027921
Covariance Phot[10]	0.000028	0.000012	0.000109
Covariance Phot[11]	-0.002826	-0.004315	0.003619
Covariance Phot[12]	-0.000745	-0.004806	0.003091
Covariance Phot[13]	0.024524	0.024521	0.024527
Covariance Phot[14]	0.021316	0.021001	0.021807
Covariance Phot[15]	0.000590	0.000589	0.000592

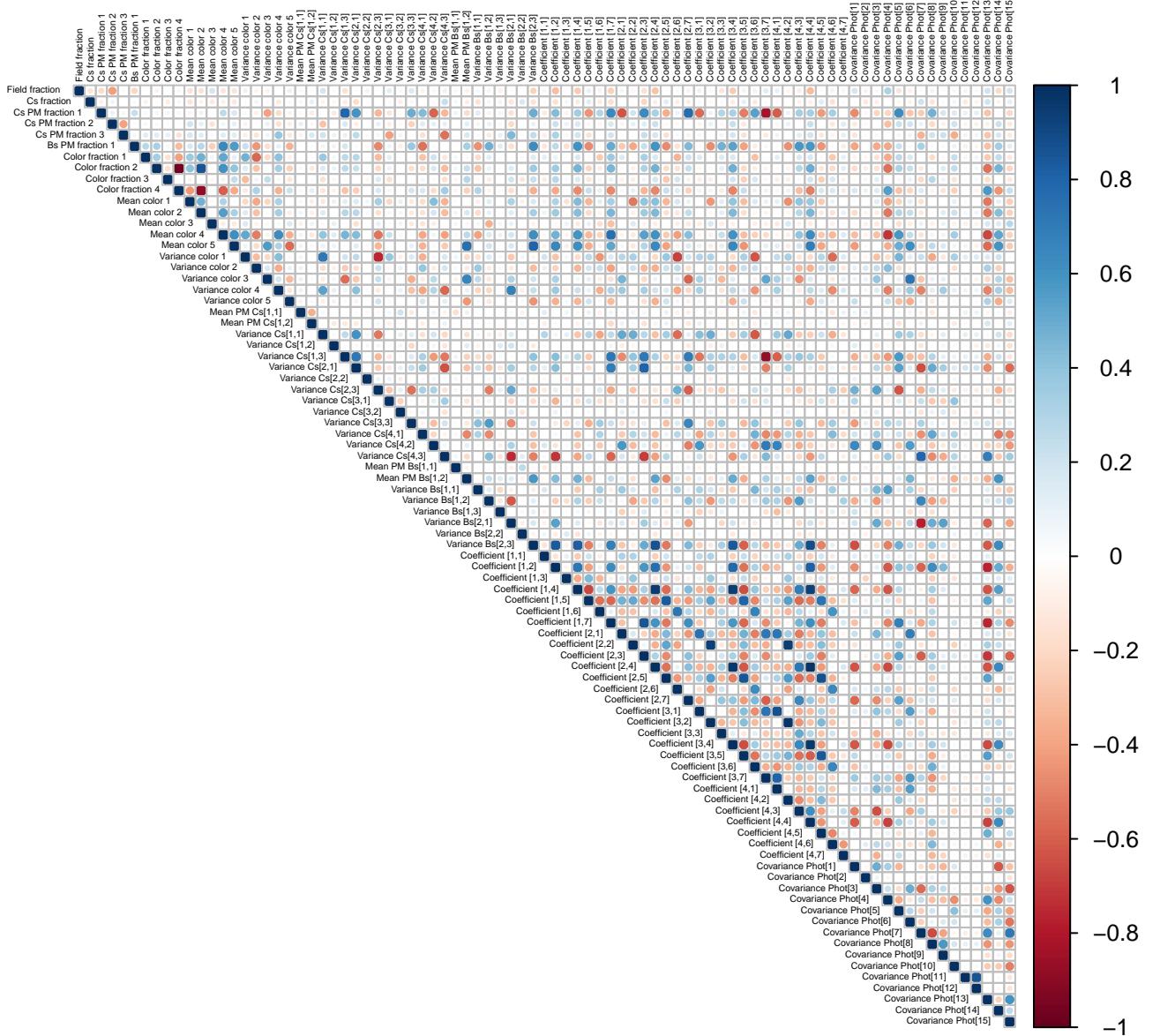


Figure 4.13: Correlation matrix of the posterior distributions of the parameters in the BHM. The colour code indicates the value of the correlation coefficient. Parameter names are the same as those in Table 4.1.

## 4.4 Updating the prior knowledge

As mentioned by Gelman (2006), the posterior distribution must be inspected to update our previous knowledge. To inspect these posterior distribution, I use the statistics reported in Table 4.1. These values indicate, for example, that the number of GMM modelling the proper motions of the single stars is overestimated. The fraction and variance of the last gaussian are both to near zero values. Probably, a better model would be that in which

the parameters of these extra gaussian will not be part of the model. Ideally, I should choose between these two models based on the evidence they show (see Section 3.1.2). To my knowledge, the only reliable approach to compute the evidence of a model inferred using MCMC, is by means of Nested Sampling (see Section 3.6.2). However, running the BHM in the *MultiNest* package lies beyond our current computational resources.

Another example of inspection of the posterior is the following. In a past run of the BHM on the RDR2, I realised that the prior distribution for the parameter modelling field fraction was too narrow. Although the maximum of the posterior distribution was allowed by this prior (by definition), the prior density at this MAP was negligible. Therefore, I updated the prior distribution to a distribution with a larger variance. Thus, I weaken the prior information.

## 4.5 Projected spatial distribution

In this Section, I present the results of the model selection analysis of the Projected Spatial Distribution (PSD) models of the Pleiades cluster. The Bayesian model selection approach and the PSD models are described in Sections 3.1.2 and 3.4, respectively. The Bayesian *evidence* is computed with uniform priors for all parameters using the Python package *PyMultiNest* (Buchner et al. 2014). It is a Python wrapper of the C++ package *MultiNest* (Feroz et al. 2009) that implements the Nested Sampling algorithm of Skilling (2004, 2006) (see Section 3.6.2).

In the Bayesian model selection methodology, the boundaries for decision making from Bayes Factors should be set *ab initio*. Thus, we discuss our results following the classical scale by Jeffreys (1961), which I show in Table 4.2.

Table 4.2: Jeffreys (1961) scale for Bayes factors.

Bayes factor	Strength of evidence
0 to 5	barely worth mentioning
5 to 10	substantial
10 to 15	strong
15 to 20	very strong
> 20	decisive

### 4.5.1 Selection of models with radial symmetry

Tables 4.3 and 4.4 summarise the evidences and Bayes factors of the radially symmetric models using the candidate members within 7 pc ( $3^\circ$ ) and 14 pc ( $6^\circ$ ), respectively.

Table 4.3: Natural logarithm of the evidence (diagonal) and Bayes factors (off-diagonal elements) for each radially symmetric model inferred from the 7 pc ( $3^\circ$ ) sample.

	King	OGKing	GKing	EFF	GP	RGP
King	-2876.2	0.13	0.04	0.03	0.04	0.41
OGKing	7.39	-2878.2	0.33	0.25	0.3	3.00
GKing	22.19	3.0	-2879.3	0.74	0.90	9.02
EFF	29.96	4.0	1.34	-2879.6	1.22	12.18
GP	24.53	3.32	1.1	0.81	-2879.4	9.97
RGP	2.45	0.33	0.14	0.08	0.1	-2877.1

Note. Bayes factors are computed with evidence of the model specified in the column header placed in the numerator and in the row header in the denominator.

Table 4.4: Natural logarithm of the evidence (diagonal) and Bayes factors (off-diagonal elements, computed as in Table 4.3) for each profile radially symmetric model inferred from the 14 pc ( $6^\circ$ ) sample.

	King	OGKing	GKing	EFF	GP	RGP
King	-5033.5	> 100	2.7	< 0.01	3.3	27.113
OGKing	< 0.1	-5023.6	< 0.01	< 0.01	< 0.01	< 0.01
GKing	0.36	> 100	-5032.5	< 0.01	1.2	10.0
EFF	> 100	> 100	> 100	-5044.1	> 100	> 100
GP	0.3	> 100	0.81	< 0.01	-5032.3	8.2
RGP	0.03	> 100	0.1	< 0.01	0.12	-5030.2

We primarily focus on the 7 pc ( $3^\circ$ ) sample, where the completeness and depth are most homogeneous (see Section 2.7). As can be seen in Table 4.3, the Bayes Factors are in the ranges between 1 and 3 (*differences barely worth mentioning* in Jeffrey’s scale) and between 3 and 10 (*substantial evidence*).

The King’s and the RGP models have the largest evidences and a Bayes Factor of 0.41 (or 2.43) between them. It implies that there is no reason to prefer one over the other. According to Jeffrey’s scale, this difference *barely worth mentioning*. At the same time, King’s model, the one with the highest evidence, shows Bayes Factors in the category between 3 and  $\sim 10$  (*substantial evidence*) or even up to 30 (*strong evidence*) when compared to the remaining models (OGKing, GKing, EFF and GP).

In Figure 4.14, we show the projections of the posterior distribution for King’s model parameters inferred from the 7 pc ( $3^\circ$ ) sample. As can be seen, the tidal radius is relatively unconstrained by the data. This can be understood in terms of the lack of observations in the vicinity of the cluster edge (tidal radius), which introduces a degeneracy. The maximum a posteriori (MAP) values of the parameters of King’s profile are  $\alpha_c = 56.65496^\circ$ ,  $\delta_c = 24.13461^\circ$ ,  $r_c = 2.0$  pc and  $r_t = 44.7$  pc. We also fit a multivariate Gaussian to the posterior distribution at the MAP estimate, it yields the following standard deviation matrix:

$$\begin{pmatrix} 0.000045 & -0.000004 & 0.000205 & -0.0008 \\ -0.000004 & 0.000038 & -0.000005 & -0.0002 \\ 0.000205 & -0.000005 & 0.20 & -0.3646 \\ -0.000805 & -0.000233 & -0.364647 & 28.048 \end{pmatrix}$$

with the above ordering of parameters ( $\alpha_c$ ,  $\delta_c$ ,  $r_c$ , and  $r_t$ ).

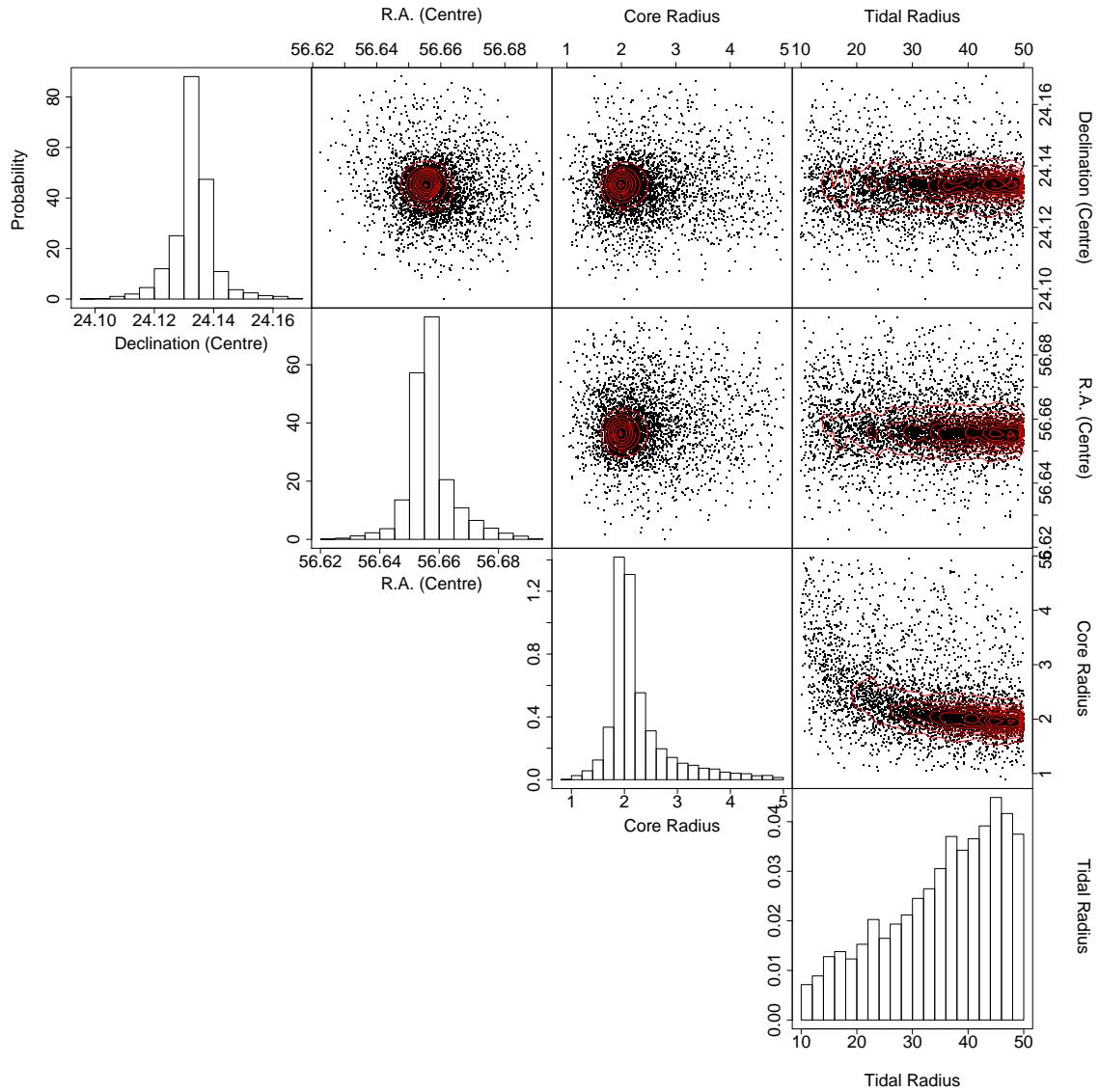


Figure 4.14: Samples of the posterior distribution of the King's profile parameters inferred from the  $3^\circ$ . The core and tidal radius are in pc.

We will first compare King's model with the GKing profile. GKing is a model with significantly less evidence and two more parameters than King's profile. The posterior distributions reveal a complete loss of interpretability of the parameters, with the core radius relatively unconstrained between 100 and 400 pc, tidal radii between 200 and 800.

The second alternative is the modified King profile with exponents fixed to the maximum evidence values (OGKing). In this case, the model has the same number of parameters as in the classical King profile, but we have selected the values of  $\alpha_1$  and  $\alpha_2$  in Eq. 3.48 that maximise the evidence (0.4 and 1.2 respectively, compared to 0.5 and 2 in the classical King profile). It is important to bear in mind that the maximum evidence is attained for the  $6^\circ$  sample, and this evidence is no longer maximum when calculated with the  $3^\circ$  sample. In this case, the parameters retain the interpretability (for obvious reasons) and the tidal radius is relatively well constrained ( $\pm 7\text{pc}$ ). However, the evidence is reduced with respect to the classical King profile despite the increased maximum likelihood. We interpret this as the result of overfitting in the process of fixing the  $a$  and  $b$  parameters. The MAP estimates of the OGKing profile parameters are  $\alpha_c = 56.65603^\circ$ ,  $\delta_c = 24.13292^\circ$ ,  $r_c = 1.15 \text{ pc}$  and  $r_t = 13 \text{ pc}$ . The standard deviation matrix of the multivariate normal fitted at the MAP is

$$\begin{pmatrix} 0.000028 & -0.000003 & 0.000009 & 0.00063 \\ -0.000003 & 0.000026 & 0.000039 & -0.00007 \\ 0.000009 & 0.000039 & 0.149843 & -0.28731 \\ 0.000627 & -0.000070 & -0.287310 & 6.69046 \end{pmatrix}.$$

The two profiles (King and OGKing) are relatively similar despite the different exponents and MAP parameter estimates.

Finally (we will not discuss the EFF or GP profiles), the RGP suffers the same loss of interpretability encountered for the GKing profile, with an unconstrained core radius in the range of hundreds of parsecs. Also the posterior distributions of the  $R_b$  and  $b$  parameters are unconstrained, with a large correlation between  $a$  and  $b$ . The resulting density profile is also cusped at the centre of the cluster.

As a summary of the results of the  $3^\circ$  sample, we find that the introduction of more flexibility in the analytical expressions of the classical radially symmetric profiles does not produce larger evidences, and results in some cases in unconstrained parameters and a loss of the interpretability associated to the original formulations. Therefore, for the data set that extends out to  $3^\circ$  there are two competing models (classical King and RGP) and there is no sufficient evidence to select amongst them. The central part is reasonably well fitted either by a King's profile with relatively well constrained parameters or with a RGP profile with large degeneracy. Only additional, perfectly acceptable prejudices like physical interpretability or the ability to compare with previous results, can be invoked to choose the King's profile over the rest.

Concerning the  $6^\circ$  sample, the OGKing profile has the largest evidence. It comes as no surprise since OGKing profile fixes the values of  $\alpha_1$  and  $\alpha_2$  to those that yield the maximum evidence. Hence, we acknowledge overfitting to the incomplete and biased sample of Pleiades members out to  $6^\circ$ . In fact, the marginal posterior distribution for the

tidal radius is sharply peaked around 16 pc, only slightly further than the observational boundary.

For the RGP model, its MAP estimates are  $\alpha_c = 56.6909785^\circ$ ,  $\delta_c = 24.13301438^\circ$ ,  $R_b = 886.84$  pc,  $\alpha = 1.44$ , and  $\beta = 70.6$ . The corresponding standard deviation matrix is

$$\begin{pmatrix} 0.0000589 & 0.000006 & -0.02 & -0.0000001 & -0.003 \\ 0.0000061 & 0.000049 & 0.01 & 0.0000134 & -0.002 \\ -0.0243964 & 0.011738 & 20173.20 & 3.0774488 & 616.625 \\ -0.0000001 & 0.000013 & 3.08 & 0.0219039 & -2.316 \\ -0.0028333 & -0.002123 & 616.63 & -2.3162244 & 331.945 \end{pmatrix}.$$

It is evident from the previous matrix that  $R_b$  and  $\beta$  are only mildly constrained. The profile is, again, characterised by a cusp in the central part of the cluster, as is the profile of the GKing alternative.

The  $6^\circ$  data set favours MAP profiles (models and MAP estimates of their parameters) with sharp peaks in the centre of the cluster. However, we caution the reader against these results. They refer to a data set that is severely biased and incomplete in ways that we cannot quantify and/or model.

The conclusion from the comparison of these radially symmetric profiles in the context of the two data sets is that i) there is no compelling reason to abandon the widely used King profile in the context of the complete and homogeneous data set out to  $3^\circ$  and ii) there are better models when evaluated on data out to  $6^\circ$ , but it remains to be found whether this is due to the completeness problems of the DANCe data in the outer regions of the Pleiades or whether they truly represent a need to make the King profile more flexible to accommodate the data. In the following we retain three of the models discussed above (King, OGKing and RGP), and take the comparison one step further in order to accommodate simple deviations from radial symmetry in the form of elliptical density contours.

### 4.5.2 Selection of models with biaxial symmetry

Concerning the models with biaxial symmetry, Table 4.5 collects the evidences and Bayes factors. We have included the results for the OGKing alternative despite the fact that we had to extend the support of the prior for the  $r_{t,a}$  and  $r_{t,b}$  parameters out to 300 pc. For smaller values of the support, the MCMC samples possess maxima at the boundaries. Even in the set up with priors extending out to 300 pc, the parameter samples lack physical interpretations with vanishingly small MAP core radii. We see that the OGKing profile is clearly to be preferred when compared against the other two models, and any of the three is to be preferred to the radially symmetric profiles of Table 4.5. However, despite the fact that the OGKing shows the best evidence and the largest Bayes factor, the unconstrained and unphysical values of its parameters allow us to discard it,

and continue exploring the luminosity segregation only in the King model.

Table 4.5: Natural logarithm of the evidences (diagonal) and Bayes factors (off-diagonal, computed as in Table 4.3) for the models with biaxial symmetry inferred from the 3° and 6° samples.

		3°			6°		
		King	OGKing	RGP	King	OGKing	RGP
3°	King	-2806.9	> 100	0.14			
	OGKing		-1527.3	< 0.01			
	RGP			-2808.9			
6°	King				-5023.8	> 100	< 0.01
	OGKing					-3255.7	< 0.01
	RGP						-5034.7

The MAP estimates of the King model on the 3° sample are  $\alpha_c = 56.66025602372^\circ$ ,  $\delta_c = 24.2038^\circ$ ,  $\phi = 2.1$  rad,  $r_{cx} = 1.84$  pc,  $r_{cy} = 1.77$  pc,  $r_{tx} = 48.9$  pc, and  $r_{ty} = 48.5$  pc. The standard deviation matrix is

$$\begin{pmatrix} 0.0124 & 0.0003 & 0.0004 & -0.0004 & 0.0002 & -0.0006 & 0.011 \\ 0.0003 & 0.0111 & -0.0025 & -0.0069 & 0.0002 & -0.0065 & 0.008 \\ 0.0004 & -0.0025 & 0.0707 & -0.0377 & 0.0540 & 0.2696 & -0.013 \\ -0.0004 & -0.0069 & -0.0377 & 0.3749 & -0.0244 & -0.0387 & -0.703 \\ 0.0002 & 0.0002 & 0.0540 & -0.0244 & 0.3722 & -0.9000 & 0.017 \\ -0.0006 & -0.0065 & 0.2696 & -0.0387 & -0.9000 & 59.3446 & -12.580 \\ 0.0113 & 0.0076 & -0.0129 & -0.7034 & 0.0174 & -12.5795 & 60.462 \end{pmatrix}$$

with the ordering of parameters used above.

Interestingly, this MAP shows no evidence of ellipticity<sup>2</sup> neither in the core radius ( $\epsilon_{rc} = 0.04$ ) nor in the tidal radius ( $\epsilon_{rt} = 0.008$ ). Despite this fact, the biaxial symmetric King models have decisive evidence when compared to the radially symmetric ones (see Table 4.7).

### 4.5.3 Selection of models with luminosity segregation

In this Section, I describe the luminosity ( $J$  band) segregation (see Section 3.4.4) included only in the radial and biaxial symmetric extensions of the King profile. As shown in the previous section, King's profile not just has large Bayes factor, but also physical interpretability. Table 4.6 summarises the evidences and Bayes Factors of the King's model with luminosity segregation in both radial and biaxial symmetric models. The parameters of these models and its evidences are inferred using the 3° and 6° samples.

Figures 4.16 and 4.17 show the posterior distributions of the parameters in the radial (King-RS) and biaxial (King-BS) symmetric King models with luminosity segregation, respectively. Both models are inferred from the complete 3° sample. Their MAP values and standard deviation matrices of their parameters are the following.

<sup>2</sup>The ellipticity is defined as  $\epsilon = 1 - \frac{b}{a}$ , with  $a$  and  $b$  the semi-major and semi-minor axis of the ellipse, respectively.

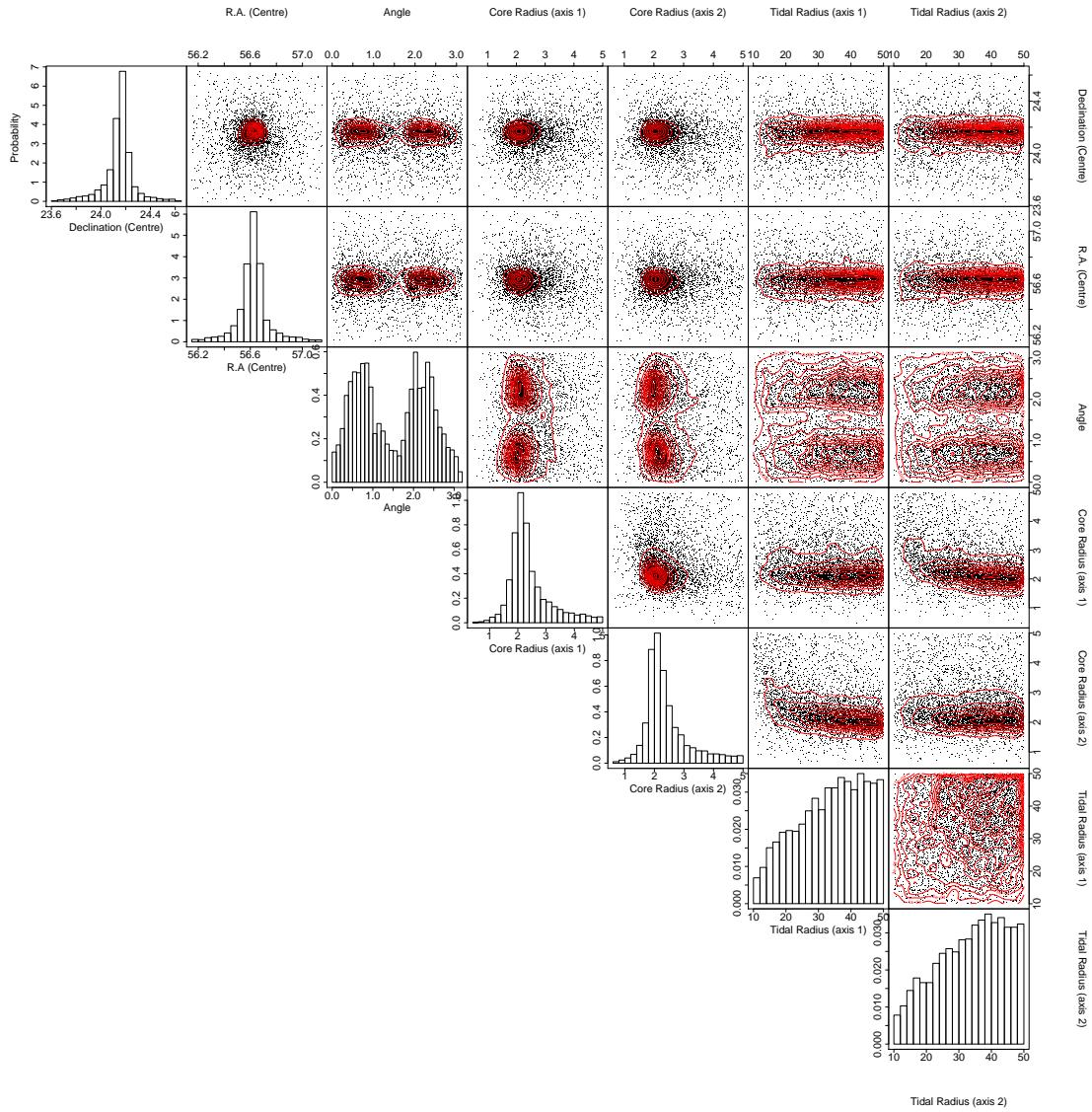


Figure 4.15: Samples from the posterior distribution of the parameters in the biaxial symmetric King model inferred from the  $3^\circ$  sample. Tidal and core radius are in units of pc.

Table 4.6: Natural logarithm of the evidences (diagonal) and Bayes factors (off-diagonal, computed as in Table 4.3) for the radial and biaxial symmetric models with luminosity segregation inferred from the  $3^\circ$  and  $6^\circ$  samples.

		$3^\circ$		$6^\circ$	
		King-Radial	King-Biaxial	King-Radial	King-Biaxial
$3^\circ$	King-Radial	-2875.5	$> 100$		
	King-Biaxial		-2803.9		
$6^\circ$	King-Radial			-5025.4	$> 100$
	King-Biaxial			-5016.4	

The MAP values of the King-RS model are  $\alpha_c = 56.6264^\circ$ ,  $\delta_c = 24.1956^\circ$ ,  $r_c = 2.115$  pc,  $r_t = 31.778$  pc, and  $\kappa = 0.127$  pc · mag $^{-1}$ . Its standard deviation matrix (with parameters in the same previous order) is

$$\begin{pmatrix} 0.0077725911 & 0.0011463608 & 0.015866753 & -0.006895138 & 0.0004781459 \\ 0.0011463608 & 0.0066805294 & 0.009640874 & 0.011311284 & 0.0002655555 \\ 0.0158667525 & 0.0096408739 & 0.380770948 & -0.844035083 & 0.0210177901 \\ -0.0068951385 & 0.0113112840 & -0.844035083 & 45.371984935 & -0.0671929448 \\ 0.0004781459 & 0.0002655555 & 0.021017790 & -0.067192945 & 0.0048742344 \end{pmatrix}$$

The MAP values of the King-BS model are  $\alpha_c = 56.6006^\circ$ ,  $\delta_c = 24.193^\circ$ ,  $\phi = 1.915$  rad,  $r_{cx} = 2.348$  pc,  $r_{cy} = 2.042$  pc,  $r_{tx} = 38.12$  pc,  $r_{ty} = 48.382$  pc and  $\kappa = 0.147$  pc · mag $^{-1}$ . Its standard deviation matrix (with parameters in the same previous order) is

$$\begin{pmatrix} 0.0139885233 & -0.0002135887 & -0.002818604 & 0.007312077 & 0.004963964 & 0.02605645 & -0.003657460 & -0.0001631951 \\ -0.0002135887 & 0.0130671231 & -0.001734756 & -0.003808827 & -0.001210257 & 0.03990854 & -0.013443550 & 0.0006731357 \\ -0.0028186038 & -0.0017347556 & 0.727100779 & 0.013468602 & 0.005179071 & -0.55449101 & 0.596346854 & 0.0011082365 \\ 0.0073120769 & -0.0038088265 & 0.013468602 & 0.422814580 & 0.054674731 & -0.19355524 & -0.679337642 & 0.0107408223 \\ 0.0049639637 & -0.0012102569 & 0.005179071 & 0.054674731 & 0.433668247 & -1.66048288 & 0.346810670 & 0.0131324789 \\ 0.0260564469 & 0.0399085427 & -0.554491013 & -0.193555244 & -1.660482881 & 89.94363663 & -13.562539506 & -0.0592770535 \\ -0.0036574599 & -0.0134435502 & 0.596346854 & -0.679337642 & 0.346810670 & -13.56253951 & 56.110848655 & -0.0089581971 \\ -0.0001631951 & 0.0006731357 & 0.001108236 & 0.010740822 & 0.013132479 & -0.05927705 & -0.008958197 & 0.0040960558 \end{pmatrix}$$

As can be seen from the MAP values and its standard deviation, the  $\kappa$  luminosity segregation parameter is similar and significative for both King-RS and King-BS models. Concerning the ellipticity, the core radius and tidal radius have large values,  $\epsilon_{rc} = 0.13$  and  $\epsilon_{rt} = 0.21$ , respectively. However, due to the large uncertainties in the tidal radii, we trust only in the ellipticity reported for the core radius. The latter value is nevertheless similar to those previously reported in the literature (0.17 and 0.1 to 0.35 reported by [Raboud & Mermilliod 1998](#); [Adams et al. 2001](#), respectively). We do not report the ellipticity value inferred from the extended 6° sample due to the incompleteness of the latter. Indeed, this ellipticity can be strongly biased by the patchy pattern of the DANCe DR2 observations.

Finally, combining the evidences and Bayes factors of the King's classical and extended models, we can conclude that there is *decisive* (in terms of Jeffrey's scale) evidence, from both the 3° and 6° data sets, supporting complex models with J-band luminosity segregation and biaxial symmetry (see Table 4.7).

The results of this Section show that, although the majority of the current models of the radial density profile of the Pleiades cluster reproduce well the observed data, only the classical King's profile and the RGP have the best evidences both in the complete 3° sample and in the extended 6° one. However, we discard the RGP profile due to its high degeneracy and unphysical values. We can perform this selection by the inclusion of restrictive priors PDFs to the physical interpretability of the models. This model selection process using non-identical priors for the competing models is explained in Section 3.1.2 (see Eq. 3.18).

Also, we have shown that the high membership probability samples, both complete (3°) and extended (6°), are better modelled, in terms of the evidence, with complex models including biaxial symmetry and luminosity segregation.

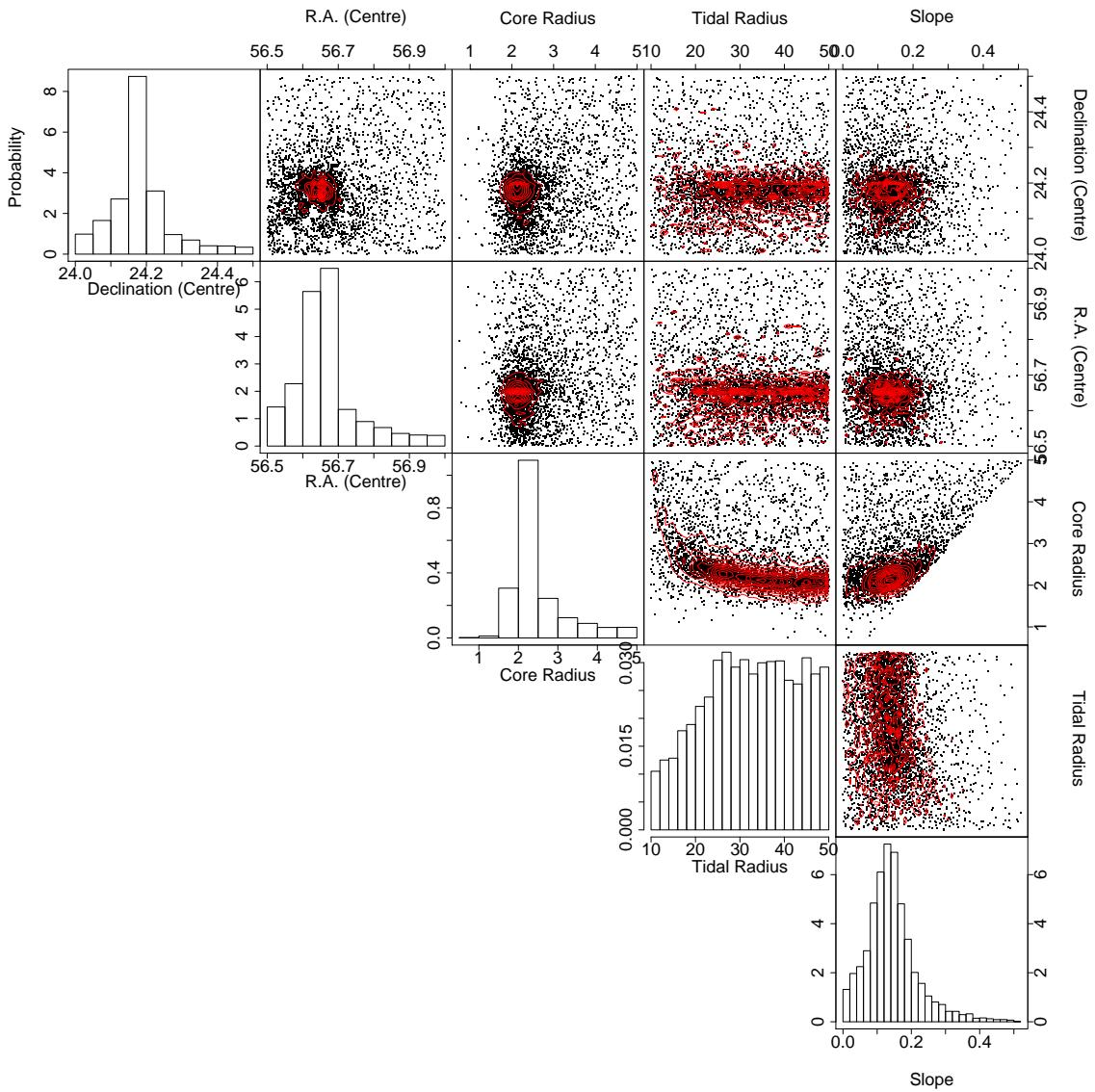


Figure 4.16: Samples of the posterior distribution of the parameters in the radially symmetric and luminosity segregated King model inferred from the  $3^\circ$  sample. The core and tidal radius are in units of pc, the slope in units of pc per magnitude.

The results of this Section establish a solid probabilistic background for: i) the classical King profile, at least in the Pleiades cluster, and ii) luminosity segregation. The latter indicates that the expected mass segregation in the Pleiades cluster (as has been indicated by previous studies, e.g. [Converse & Stahler 2010](#)) is supported by Bayesian evidence.

## 4.6 Proper motions distribution

The bivariate proper motions distributions of both single and EMB is directly recovered by the BHM by means of the posterior distributions of the parameters in the GMM modelling these populations. These bivariate distributions are depicted in Fig. 4.18.

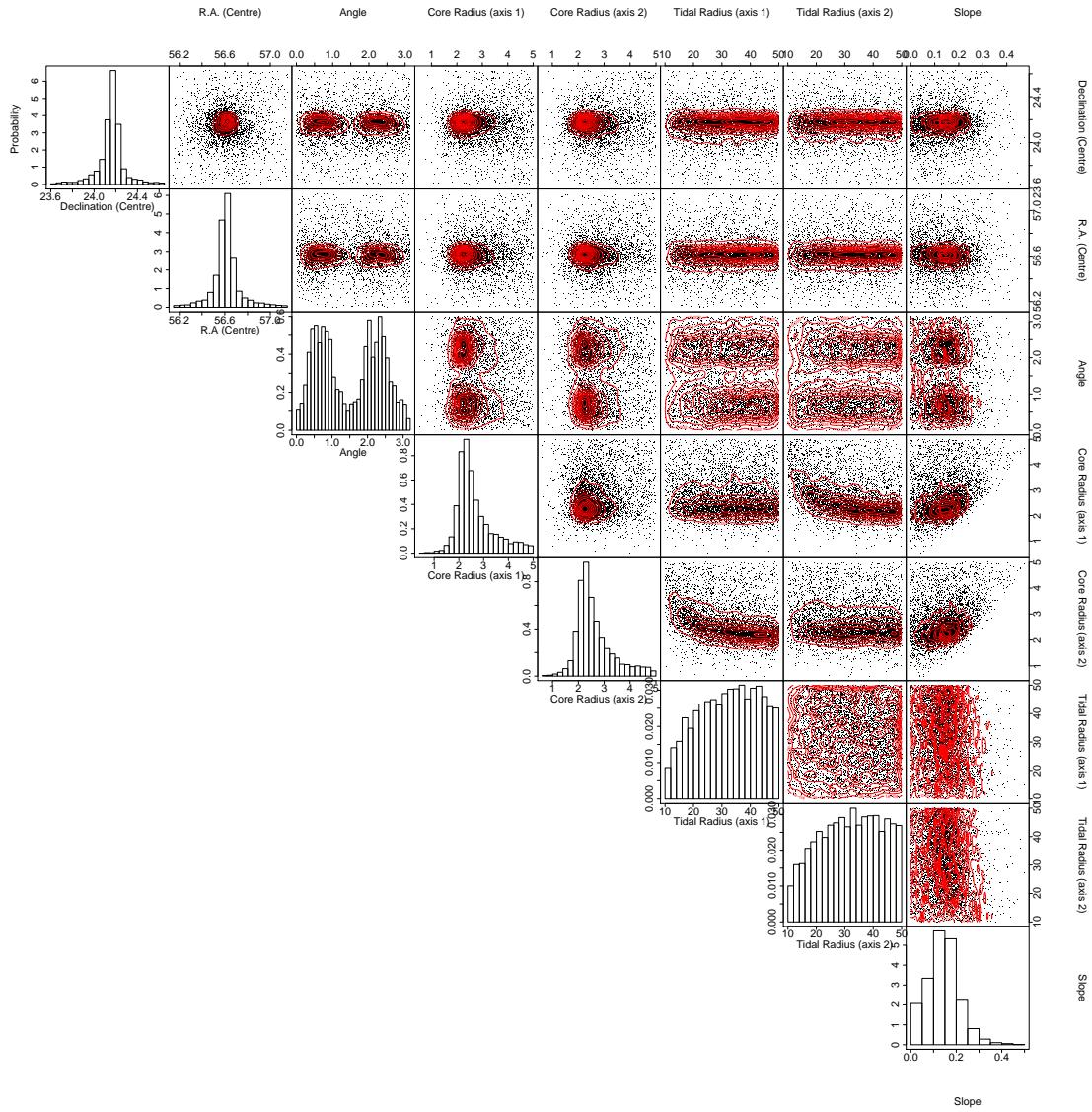


Figure 4.17: Samples of the posterior distribution of the parameters in the biaxial symmetric and luminosity segregated King model inferred from the  $3^\circ$  sample. The core and tidal radius are in units of pc, the slope in units of pc per magnitude.

The univariate projections of these distributions, in the  $\mu_\alpha \cdot \cos(\delta)$  and  $\mu_\delta$  components, are shown in Figs. 4.19 and 4.20. Also, I show in this figures the densities in the same proper motions projections of the candidate members of Bouy et al. (2015), and those found by BHM, which have membership probabilities grater than 0.84. Interestingly, the densities rendered by both samples of members are almost identical. Nevertheless, the density resulting from the sample of the posterior distributions of the parameters in the GMMs, differ from the kernel density estimations of the corresponding populations. This difference results from the distinct populations of the two distributions. The BHM finds the posterior distribution of its parameters using the likelihood of all the objects in the data set. The contribution that individual objects have to the total cluster likelihood

Table 4.7: Natural logarithm of the evidences (diagonal) and Bayes factors (off-diagonal, computed as in Table 4.3) for King models with radial symmetry (King-R), biaxial symmetry (King-B), and radial and biaxial symmetry with luminosity segregation (King-RS and King-BS, respectively). These evidences are computed using the samples of  $3^\circ$  and  $6^\circ$ .

		$3^\circ$				$6^\circ$			
		King-R	King-B	King-RS	King-BS	King-R	King-B	King-RS	King-BS
$3^\circ$	King-R	-2876.2	$> 100$	2.01	$> 100$				
	King-B		-2806.9	$< 0.1$	20.1				
	King-RS			-2875.5	$> 100$				
	King-BS				-2803.9				
$6^\circ$	King-R					-5033.5	$> 100$	$> 100$	$> 100$
	King-B					-5023.8	0.2	$> 100$	
	King-RS						-5025.4	$> 100$	
	King-BS								-5016.4

can be thought to be proportional to their cluster membership probability. Therefore, the observed difference results from those objects whose cluster membership probability is lower than 0.84. The expected number of cluster members in the probability range 0-0.84 amounts to the non negligible value of  $\sim 1100$  objects. These objects contaminate the posterior distribution of the parameters in the BHM with an expected value of 5.3% (see Section 4.1). Nevertheless, within them also lies a 10% of true cluster members (as estimated in Section 4.1). Discarding these true cluster members from the statistical analysis will render it biased. Instead, we decided to include these 10% possible true cluster members at the small price of a 5.3% of contamination.

## 4.7 Luminosity distribution

This Section describes the process to obtain the  $J, H$ , and  $K_s$  absolute magnitude distributions from the posterior distributions of the parameters in the BHM. Later, I compare them to those found by Bouy et al. (2015), in the  $K_s$  band specifically. As in the previous section, I also compare these distributions with those resulting from the kernel density estimates of the candidate members of the BHM.

### 4.7.1 Derivation of the magnitude distributions

To derive the  $J, H, K_s$  magnitude distributions, I first transform the true  $CI$  distribution into the  $J, H, K_s$  apparent magnitude distributions. In this transformation, the fractions of EMB must be taken into account. To do this, the magnitude distributions of both single and EMB populations are mixed according to their fractions. In the following paragraphs, I describe the process to obtain the  $K_s$  apparent magnitude. This process is similar for the rest of the bands.

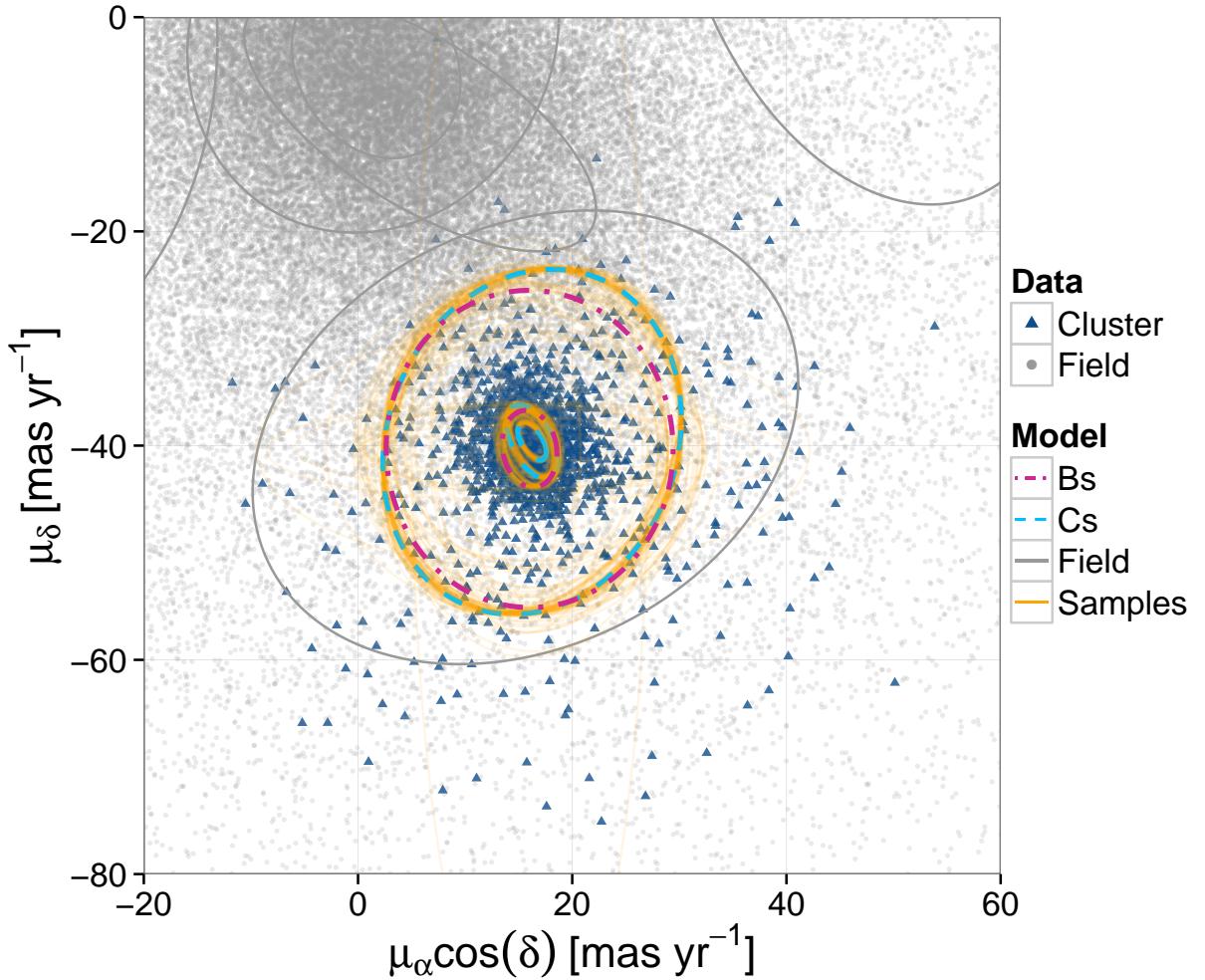


Figure 4.18: Proper motion distributions recovered by the BHM. The dashed and dot-dashed ellipses represent the mode of 100 samples (orange lines) from the posterior covariance matrices in the cluster and EMB GMMs, respectively. Reproduced from Figure 8 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, Vol. submitted.

Since we aim at the probability distribution of  $K_s$ , and it depends on the true  $CI$ , I use this as a nuisance parameter that I later marginalise. Thus,

$$p(K_s|\boldsymbol{\theta}_c) = \int p(K_s, CI|\boldsymbol{\theta}_c) \cdot dCI = \int p(K_s|CI, \boldsymbol{\theta}_c) \cdot p(CI|\boldsymbol{\theta}_c) \cdot dCI.$$

The term  $p(K_s|CI, \boldsymbol{\theta}_c)$  represents the probability of  $K_s$  given the true  $CI$  and the cluster parameters  $\boldsymbol{\theta}_c$ . It is given by Eq. 3.31. The second term,  $p(CI|\boldsymbol{\theta}_c)$  corresponds to

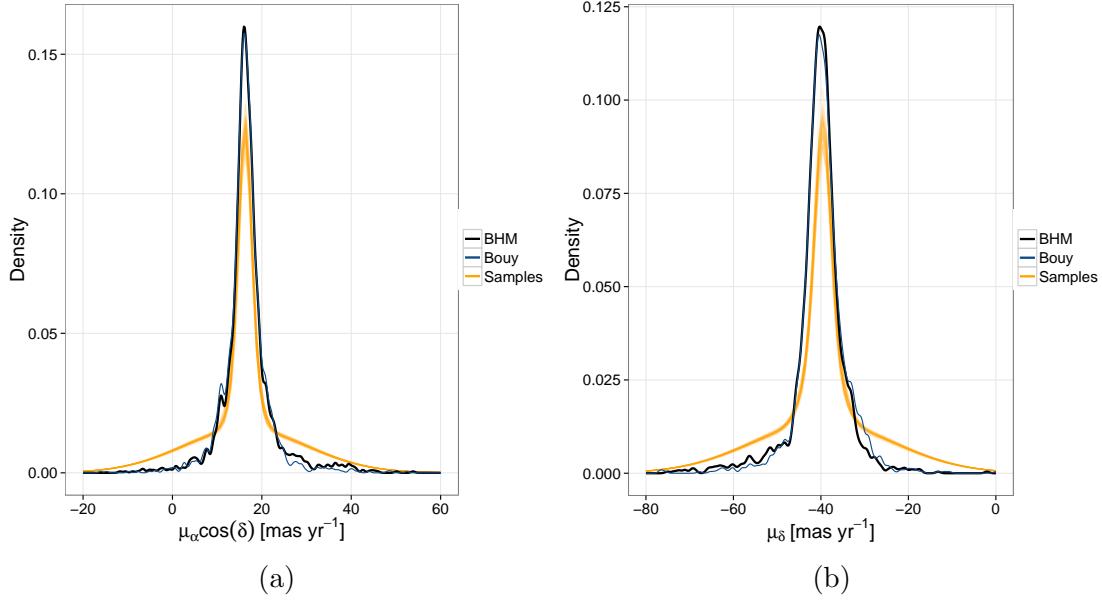


Figure 4.19: Proper motions densities resulting from: a 100 element sample from the posterior distributions of parameters in the GMM modelling the single stars (orange spaghetti lines), the kernel density estimation of the BHM candidate members classified as single stars (those whose cluster membership probability is higher than 0.84 and EMB membership probability is lower than 0.5), and, the kernel density estimation of the candidate members of Bouy et al. (2015) whose photometry lies below the EMB sequence (blue line).

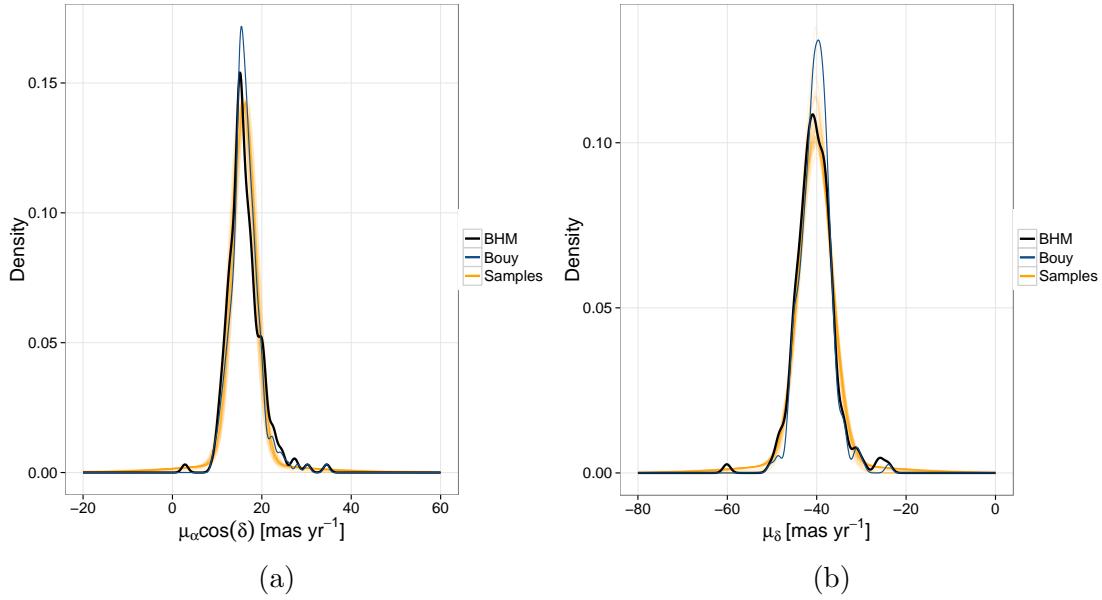


Figure 4.20: Proper motions densities resulting from: a 100 element sample from the posterior distributions of parameters in the GMM modelling the EMB stars (orange spaghetti lines), the kernel density estimation of the BHM candidate members classified as EMB stars (those whose cluster membership probability is higher than 0.84 and EMB membership probability is higher than 0.5), and, the kernel density estimation of the candidate members of Bouy et al. (2015) whose photometry lies near the EMB sequence (blue line).

the GMM modelling the distribution of the true  $CI$ , it is given by Eq. 3.32.

We include the EMB distribution with an amplitude equal to their fraction,  $1 - \pi_{CB}$ . Thus,

$$\begin{aligned}
p(K_s|\boldsymbol{\theta}_c) &= \int [\pi_{CB} \cdot p_{Cs}(K_s|CI, \boldsymbol{\theta}_c) + (1 - \pi_{CB}) \cdot p_{Bs}(K_s|CI, \boldsymbol{\theta}_c)] \\
&\quad \cdot p(CI|\boldsymbol{\theta}_c) \cdot dCI \\
&= \pi_{CB} \int p_{Cs}(K_s|CI, \boldsymbol{\theta}_c) \cdot p(CI|\boldsymbol{\theta}_c) dCI \\
&\quad + (1 - \pi_{CB}) \int p_{Bs}(K_s|CI, \boldsymbol{\theta}_c) \cdot p(CI|\boldsymbol{\theta}_c) \cdot dCI.
\end{aligned} \tag{4.1}$$

In this equation,  $Cs$  and  $Bs$  are the subindices used to distinguish the probability of  $K_s$  under the cluster and EMB photometric models, respectively. These probabilities are defined for the vector of photometric measurements,  $\mathbf{d}_{ph}$  (see Eq. 3.31). Since we are interested only in the distribution of  $K_s$  (by now), we marginalise the rest of the photometric entries, including the observed  $CI$  (I use a tilde over the observed quantities). Also, the integration limits must change to those of the truncated true colour distribution ( $CI_{min} = 0.8, CI_{max} = 8$ ). Hence,

$$\begin{aligned}
p(K_s|\boldsymbol{\theta}_c) &= \pi_{CB} \int_{CI_{min}}^{CI_{max}} \left[ \left[ \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \right] \right. \\
&\quad \cdot \left. \int_{\tilde{CI}, \tilde{Y}, \tilde{J}, \tilde{H}} \mathcal{N}(\{\tilde{CI}, \tilde{Y}, \tilde{J}, \tilde{H}, K_s\} | \mathcal{S}(CI, \boldsymbol{\beta}), \Sigma_{clus}) d\tilde{CI} d\tilde{Y} d\tilde{J} d\tilde{H} \right] \cdot dCI \\
&\quad + (1 - \pi_{CB}) \int_{CI_{min}}^{CI_{max}} \left[ \left[ \sum_{i=1}^5 \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \right] \right. \\
&\quad \cdot \left. \int_{\tilde{CI}, \tilde{Y}, \tilde{J}, \tilde{H}} \mathcal{N}(\{\tilde{CI}, \tilde{Y}, \tilde{J}, \tilde{H}, K_s\} | T_{Bs}(\mathcal{S}(CI, \boldsymbol{\beta})), \Sigma_{clus}) d\tilde{CI} d\tilde{Y} d\tilde{J} d\tilde{H} \right] \cdot dCI.
\end{aligned}$$

The derivations of the  $J$  and  $H$  magnitude distributions are similar. Since this process takes into account the unresolved EMB and the so called single stars, which in fact could be binaries with low mass ratios, then I call these distributions the apparent system magnitude distributions.

The previous distributions, together with the parallax and extinction of the cluster, are used to obtain the luminosity distributions, more properly the absolute system magnitude distributions. I assume that the distribution of parallaxes of the Pleiades members is normally distributed with mean, 7.44 mas, and standard deviation 0.42 mas (Galli et al. 2017). Then, to obtain the absolute magnitude distributions, I subtract this parallax

distribution, by means of a convolution, to the  $J, H, K_s$  magnitude distributions. Finally, I deredden the previous distributions employing the canonical value of extinction for the Pleiades:  $A_v = 0.12$  mag ([Guthrie 1987](#)). This last values were transformed into the  $J, H, K_s$  extinctions using the extinction law of [Cardelli et al. \(1989\)](#).

The completeness limits of the derived luminosity distributions are acquired as follows. Since the BHM methodology prescribes the *true* photometric quantities based on the *true* colour index  $CI$ . Then, the completeness limits of the  $CI$  distribution dictate those of the luminosity distributions.

[Bouy et al. \(2015\)](#) mention that, due to the heterogeneous origins of the DANCe DR2 survey, the spatial coverage and sensitivity of the survey is also not homogeneous. To remedy this issue, they identify a region with complete spatial coverage, the inner three degrees of the cluster (see Fig. 2.12). Then, they restricted their photometric analysis to this spatially complete region.

Restricting their sample of candidate members to this inner region results in a sample of candidate members that is spatially biased. If any dynamical process has been set on the cluster such that the mass distribution of its members is not uniformly distributed in the space, then a spatial cut in a sample of candidate members will result in a bias on the mass distribution. One of such dynamical process is the mass segregation, which, as suggested by [Adams et al. \(2001\)](#) may have happen in the Pleiades cluster. As proven in Section 4.5.3, there is strong evidence in favour of luminosity segregation in the Pleiades, at least in the  $J$  band.

Thus, to avoid such biases, I assume that the UKIDSS survey, which is the most profound from among the contributions to the DANCe DR2 catalogue, provides the homogeneous spatial and sensitivity coverage at faint magnitudes (see Fig. 2.12). This survey thus provides the upper completeness limits, which are essentially those reported in the Appendix A of [Bouy et al. \(2015\)](#),  $i \sim 23$  mag and  $K_s \sim 18$  mag.

Since we are interested in the completeness limits of the  $CI$ , and this equals  $i - K_s$ , then its completeness limits are defined by those of  $i$  and  $K_s$  together. In Fig. 4.21, I show the  $K_s$  and  $i$  kernel density estimate computed using all sources in the Pleiades DANCe DR2. As can be seen from this Figure, the point with maximum density, which corresponds to  $i = 21.4$  mag and  $K_s = 18.1$  mag, should be used to set the upper completeness limits. Notice that, due to the use of the two dimensional density, the upper completeness limit in  $i$  is reduced with respect to that of the univariate  $i$  distribution. This density shows a sharp decline at bright magnitudes, probably due to saturation of the detectors. To be conservative, I choose  $i = 13.2$  mag and  $K_s = 11.0$  mag as the lower completeness limits.

The  $CI$  completeness interval is then defined as that of all the points, along the cluster sequence in the  $K_s$  vs.  $i - K_s$  CMD, for which  $i$  and  $K_s$  are bounded by their upper and lower completeness limits, respectively. This results in a completeness interval of  $2.7 < CI < 5.6$  mag. With it, and the cluster sequence (the splines), I derive the

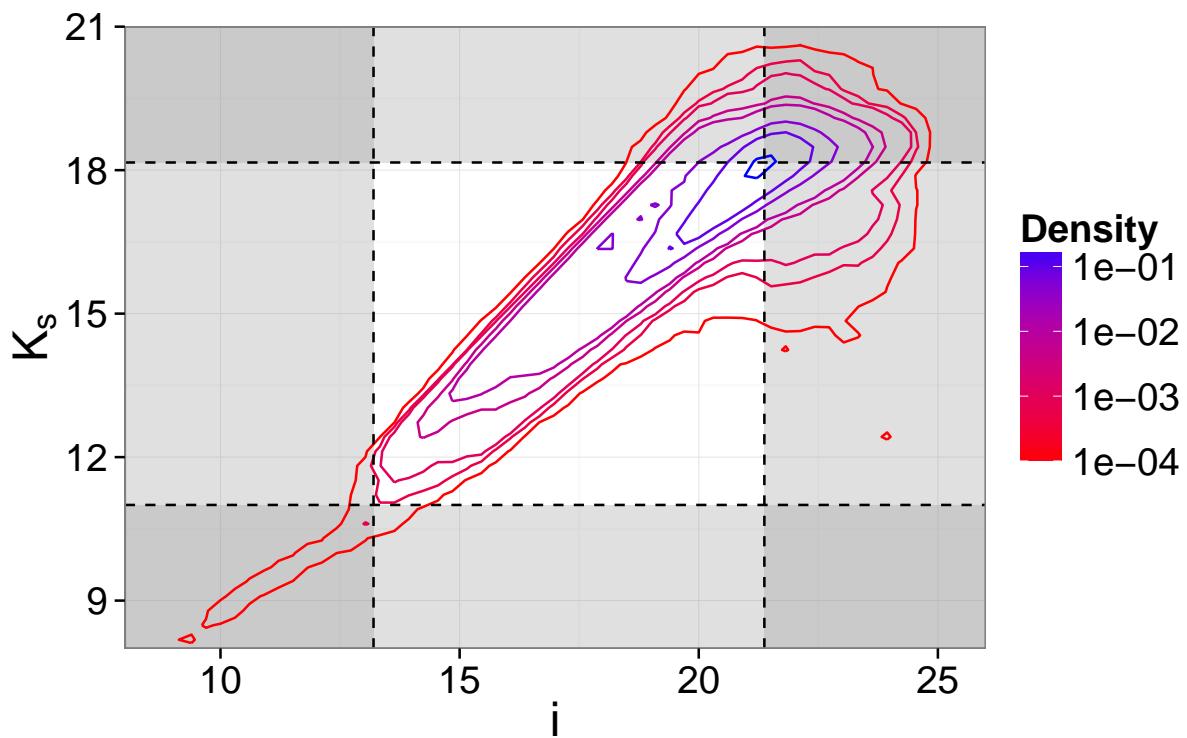


Figure 4.21: Density of all DANCE DR2 sources in  $K_s$  and  $i$  magnitudes. Lines show the chosen completeness limits,  $13.2 < i < 21.4$  mag and  $11 < K_s < 18.1$  mag. The grey area is considered incomplete. Reproduced from Figure 9 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, A&A, Vol. submitted.

completeness intervals for the  $J, H, K_s$  bands. Finally, I transform these intervals to absolute magnitudes and deredden them.

The luminosity distributions of the  $J, H, K_s$  bands, together with their completeness limits are shown in Fig. 4.22. I call them model BHM. For the sake of comparison, I also show the following luminosity distributions. First, the luminosity distributions of objects classified as candidate members in the BHM. I call these the discrete BHM distributions. Also, I plot the luminosity distribution resulting from the candidate members of [Bouy et al. \(2015\)](#), I call them the discrete Bouy. Since the discrete luminosity distributions, both Bouy and BHM, rely on the magnitudes of the individual candidate members, and many of these objects have missing value entries, then I impute their missing entries using those of the nearest euclidean neighbour.

The difference between the luminosity distributions derived using the posterior distribution of the parameters in the BHM and discrete BHM distributions, comes as well from the fact that the candidate members are not a random sample of the cluster population. As explained before, the luminosity distributions derived from the parameters in the model take into account all objects proportionally to their cluster membership probability. The discrete BHM uses only the high probability candidate members, thus they may be biased by the probability threshold. In addition, the missing value entries of all the objects in the model BHM distributions are marginalised, while in the discrete BHM they are imputed.

On the other hand, the differences between the discrete distributions, the BHM and that of [Bouy et al. \(2015\)](#), arise mainly at the bright and faint ends ( $K_s \sim 4$  mag and  $K_s \sim 11$  mag). I hypothesise that the origin of these differences lie in the different list of candidate members. As it has been discussed, the model of [Bouy et al. \(2015\)](#) is constructed only based in fully observed objects. The regions where the objects with missing entries happen more frequently is in the bright and faint regions. Therefore, the observed differences in the luminosity distributions may arise from the incomplete treatment that those authors made of objects with missing entries.

## 4.8 Mass distribution

In this Section I describe the procedure to transform the luminosities distributions into mass distributions. Transforming a probability distributions requires the transformation *per se* and its derivative (see Section 3.1). Once the mass distributions is obtained, then, I compare it to the Initial Mass Functions (IMFs) of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#). Finally, I conclude this section with the analysis of some simple toy models that can be fitted to the derived mass distribution.

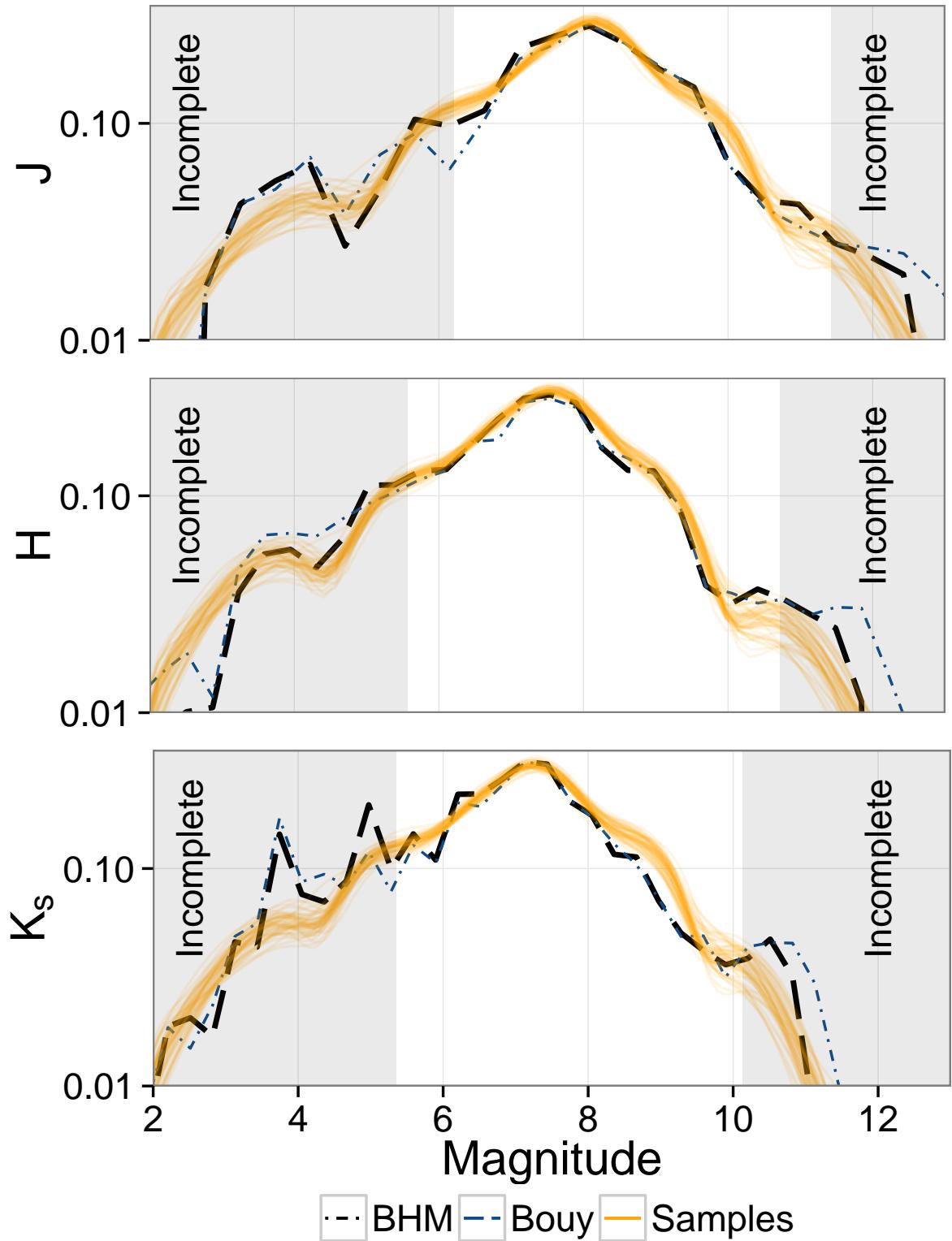


Figure 4.22: Luminosity distribution of  $J$ ,  $H$ ,  $K_s$  bands (orange spaghetti lines). Also shown: the regions of incompleteness, and, the luminosity distributions computed from: the candidate members of Bouy et al. (2015) (dot-dashed blue line), and our candidate members, ( $p_{84\%} > p_t$ , dashed black line). Reproduced from Figure 10 of Olivares et al. (2017), *The Seven sisters DANCE III: Hierarchical Bayesian model*, submitted submitted.

### 4.8.1 The mass-luminosity relation

The mass-luminosity relation is the non linear transformation that enables us to obtain the mass distribution from the luminosity distributions. Given the values of the upper limits of the luminosity distributions (the faint ends), the mass-luminosity relation relies entirely on the current models of stellar atmospheres. Among the different flavours of theoretical stellar evolution models in the literature (those from the Pisa, Padova, Trieste, Geneva, and Lyon research groups) we choose the BT-Settl models of [Allard et al. \(2012\)](#). These models go deeper into the lower masses reaching the planetary mass range thus allowing a complete coverage of our luminosity distributions. The rest of the models stay in the  $0.1 - 10 M_{\odot}$  range, with the *PARSEC* models being the ones reaching the  $0.1 M_{\odot}$  limit ([Bressan et al. 2012](#)).

From the BT-Settle grid I choose the CIFIST2011bc for the 2MASS AB photometric system, 120 Myr and solar metallicity. I choose this photometric system because it covers the dynamic range of the DANCE DR2 survey. The age and metallicity values are the closest, within the grid, to the canonical ones (see Section 3.2.1). This grid returns values of the luminosity for certain non uniformly distributed values of the mass. As shown in Eq. 3.6, the transformation of a probability distribution, in this case the luminosities, into the mass distribution is proportional to the derivative of the transformation, which must be continuous. To avoid the discontinuities in the derivatives produced by the grid, we fit the values from the grid using spline series (see Fig. 4.23a). Then, derivative is obtained from these continuous series (see Fig. 4.23b). It is important to notice the following two assumptions. First, I assume that the luminosity distributions in  $J$ ,  $H$  and  $K_s$  bands are independent between them, and then I obtain a mass distribution for each one of them. Second, I assume that the transformation from luminosities to masses does not have any associated uncertainty. These assumptions must be taken since the isochrone models do not provide neither uncertainties nor a way to incorporate correlations between the mass distributions of distinct photometric bands.

Figure 4.23a shows the spline fit to the mass-luminosity relations of the BT-Settl absolute  $J$ ,  $H$  and  $K_s$  magnitudes (black points) as a function of the mass. Figure 4.23b shows the derivative of mass-luminosity relation. The grey shaded areas represent the incompleteness regions of the DANCE survey (see previous section).

### 4.8.2 Present day system mass distribution

The mass distribution is independently obtained for the  $J$ ,  $H$ ,  $K_s$  luminosity distributions by means of the mass-luminosity relations described in the previous section. Since the luminosity functions of Sect. 4.7 correspond to the luminosity of systems (single stars unresolved binaries and multiple systems), then, the derived mass function corresponds to the Present Day System Mass Distribution (PDSMD). Figure 4.24 shows the logarithmic

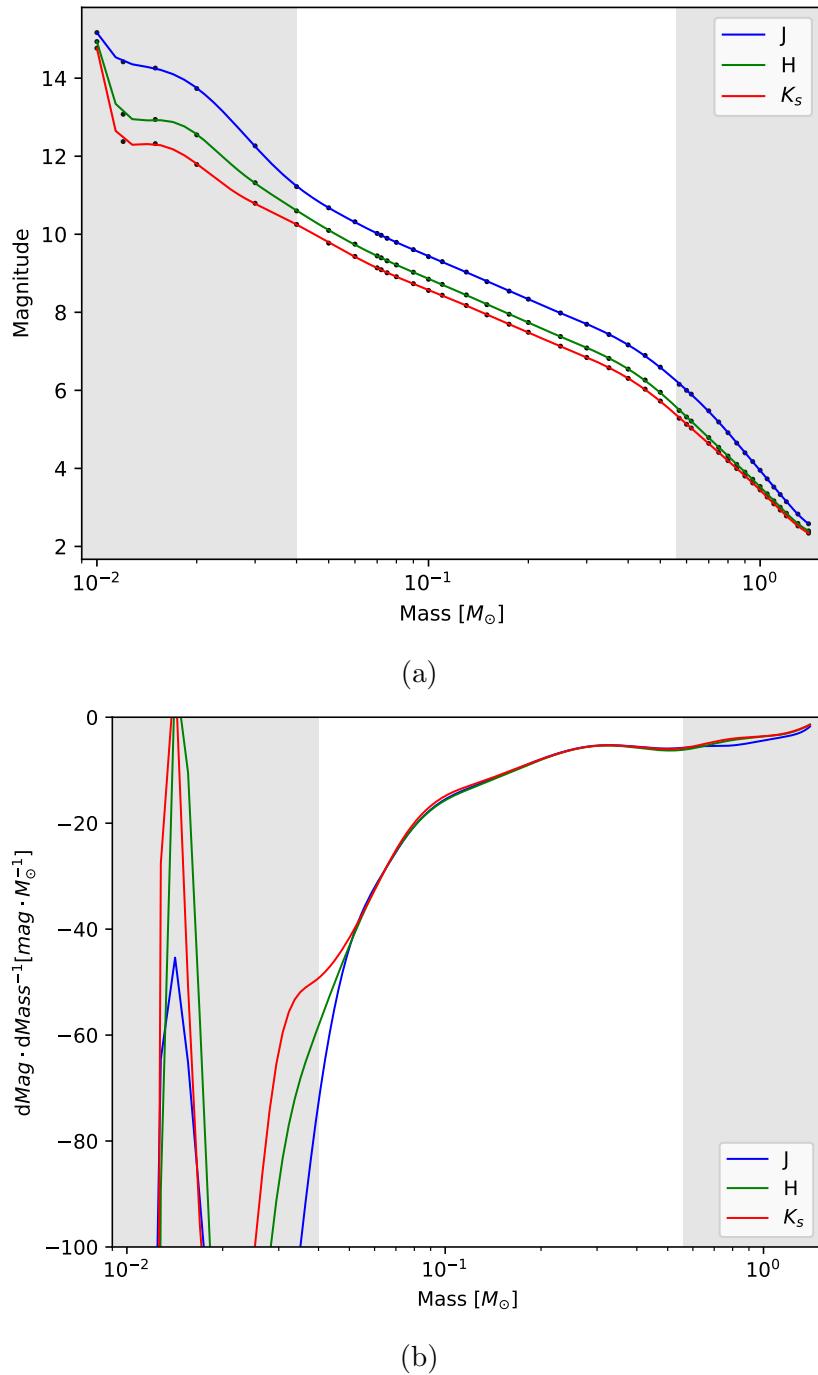


Figure 4.23: Upper panel: Mass-luminosity relations from the BT-Settl models (Allard et al. 2012) for the  $J$ ,  $H$  and  $K_s$  bands of the 2MASS photometric system (black dots). Also shown, the splines fitted to the previous relations. Bottom panel: The derivative of the mass-luminosity relations in the upper panel. The incompleteness regions of the DANCe survey (grey areas) are shown in both panels.

PDSMD ( $\xi_L$ ) for the  $J$ ,  $H$ ,  $K_s$  bands normalised on the completeness limits of the DANCe survey. The logarithmic representation of the mass distribution is a transformation from the natural variable of mass into the logarithm of 10 scale. It is customary to represent the mass distribution in this scale.

For the sake of comparison, Figure 4.25 shows the PDSMD ( $\xi_L$ ) for the  $K_s$  band of the

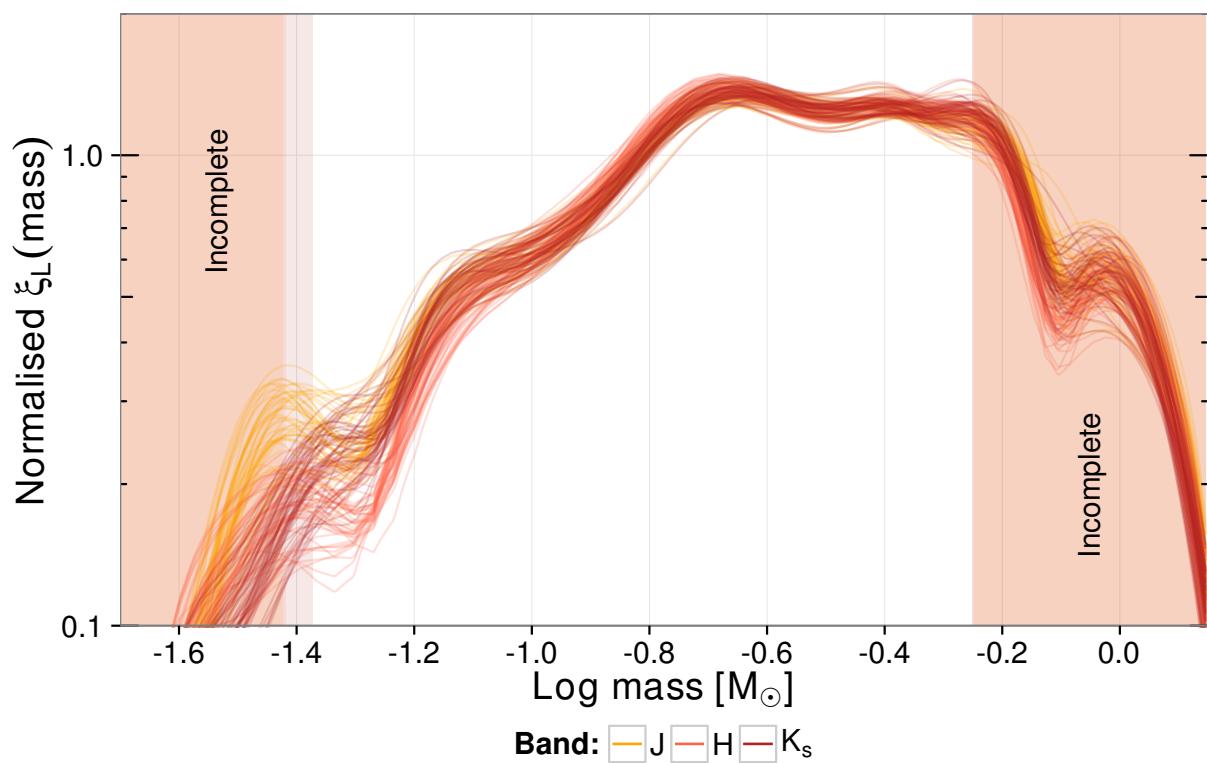


Figure 4.24: Normalised logarithmic PDSMD in  $J, H, K_s$  band. Also shown the completeness limits computed in previous section and transformed with the mass-luminosity relation.

previous Figure, together with the three-slope power-law function of [Bouy et al. \(2015\)](#), and the Initial Mass Functions (IMF) of [Thies & Kroupa \(2007\)](#) and [Chabrier \(2005\)](#). The standard uncertainties in [Chabrier \(2005\)](#) IMF are those reported in [Chabrier \(2003a\)](#).

This Figure shows that the PDSMDs derived from the BHM compare well, at least in the completeness interval, with the one proposed by [Bouy et al. \(2015\)](#). The discrepancies between these two, above  $0.3M_{\odot} (-0.5 < \log M/M_{\odot})$  particularly, may have its origin on the following aspects.

The PDSMF of [Bouy et al. \(2015\)](#) is computed using only their candidate members within the central three degree region of the Pleiades DANCe DR2. First, their list of candidate members is not the same as those found by the BHM. Second, the PDSMD derived from the BHM uses all objects in the data set, not just the high membership probability candidates. Third, as mentioned in Section 4.7, the cut to the central three degree region may have biased the derived PDSMF of [Bouy et al. \(2015\)](#). Therefore, the lack of objects that it shows, in the mass range  $0.3 - 0.7M_{\odot} (-0.5 < \log M/M_{\odot} < -0.2)$  particularly, may has its origin in the objects that [Bouy et al. \(2015\)](#) did not included in his analysis: those lying outside the inner three degree region.

For the sake of completeness, I fit a simple model to the PDSMD of the BHM. To do it, I proceed as follows. First, I select three competing models: a log-normal function (like that of [Chabrier \(2003a, 2005\)](#)), and two power-law functions of the form  $m^{-\alpha}$  with two and three power-law segments. Second, from a 100 sample distribution of the PDSMD in the  $K_s$  band (the ones shown as spaghetti lines in Fig. 4.25) I compute the mean distribution. This is, at each of the 200 points grid spanning the completeness interval, I obtain the mean of the probabilities given by the 100 distributions. With this mean distribution I draw a  $10^4$  synthetic sample of masses. Third, using *PyMultiNest* ([Buchner et al. 2014](#)), I infer the parameters of the three models given the masses of the synthetic sample. Table 4.8 gives the MAP of the parameters in each model together with the natural logarithm of the evidences (see Sect. 3.1.2). Judging by [Jeffreys \(1961\)](#) scale of evidence (see Table 4.2), there is decisive evidence supporting the two and three segment power-law models. However, the Bayes factor of the latter indicates a difference in evidence which barely worth mentioning. Thus, in the following I compare with the two-segment power-law model that is shown on Fig. 4.25 (as the black solid line).

The two segment power-law model agrees with the three segment model of [Bouy et al. \(2015\)](#). However, there are still differences, which are clear at the low and high mass ends particularly. Nevertheless, it is in clear discrepancy with the IMFs of [Chabrier \(2005\)](#), ( $m_c = 0.25^{+0.021}_{-0.016}$  and  $\sigma = 0.55^{+0.05}_{-0.01}$ , the uncertainties are those reported by [Chabrier 2003a](#), for single objects) and of [Thies & Kroupa \(2007\)](#).

The discrepancy between the IMFs and the PDSMD derived from the BHM and the PDSMF of [Bouy et al. \(2015\)](#) may have its origin on the not yet established uncertainties in the mass-luminosity relation, on dynamical effects associated with age, or in a combination

of the previous. In the next section I compare the PDSMD of the Pleiades with that of other younger and older clusters in order to analyse if there is evidence of dynamical effects associated with age.

Table 4.8: Parameters and evidence of models fitted to the PDSMD

Model	Parameters	Log Evidence
LogNormal	$m_c = 0.36 \pm 0.03$ $\sigma = 0.46 \pm 0.02$	$18.1 \pm 0.1$
Two Segments	$\alpha_0 = -0.11 \pm 0.06$ $m \in [0.04, 0.22 \pm 0.01]$ $\alpha_1 = 1.13 \pm 0.1$ $m \in [0.22 \pm 0.01, 0.56]$	$2222.7 \pm 0.4$
Three Segments	$\alpha_0 = -0.05 \pm 0.6$ $m \in [0.04, 0.08 \pm 0.03]$ $\alpha_1 = -0.1 \pm 0.1$ $m \in [0.08 \pm 0.03, 0.22 \pm 0.01]$ $\alpha_2 = 1.13 \pm 0.1$ $m \in [0.22 \pm 0.01, 0.56]$	$2221.2 \pm 0.3$

However, ending this section, I use the PDSMD to give a lower limit to the mass of the cluster. Since the RDR2 data set still lacks the very low mass range and most of the high mass range, the mass derived from this PDSMD is only a lower limit to the mass of the cluster. From the PDSMD, the cluster mean mass in the entire mass range is  $0.257 \pm 0.006 M_\odot$ . Thus, the product of this mean mass with the expected number <sup>3</sup> of cluster members ( $3301 \pm 140$ ), gives the expected mass of the cluster in this mass range. This value is  $845^{+38}_{-33} M_\odot$ .

Finally, I notice that, as mentioned in Sect. 4.8.1, the uncertainties in the mass-luminosity relations are yet to be established. Thus the quoted uncertainties of our mass results are underestimated.

## 4.9 The mass distribution on time

As I mention in the previous section, the observed differences between the present day mass distribution and the initial mass functions may have their origin on the temporal evolution of the cluster population. To test this hypothesis, I compare the Pleiades PDSMD ( $\sim 125$  Myr) with those of the younger Trapezium ( $\sim 1$  Myr) and Hyades ( $\sim 600$  Myr) PDMD. These can be thought as snapshots of the Pleiades past and future mass distributions.

Although this comparison formally lies beyond the objectives of the present work, nevertheless, it gives an idea of the importance that the PDSMD of other NYC have in the understanding of the formation and evolution of the mass distribution.

Figure 4.26 shows the PDSMD from the Pleiades, together with those of the Trapezium

---

<sup>3</sup>As explained before, the expected number of cluster members is the integral, over the whole range of membership probabilities, of number of objects at each membership probability value.

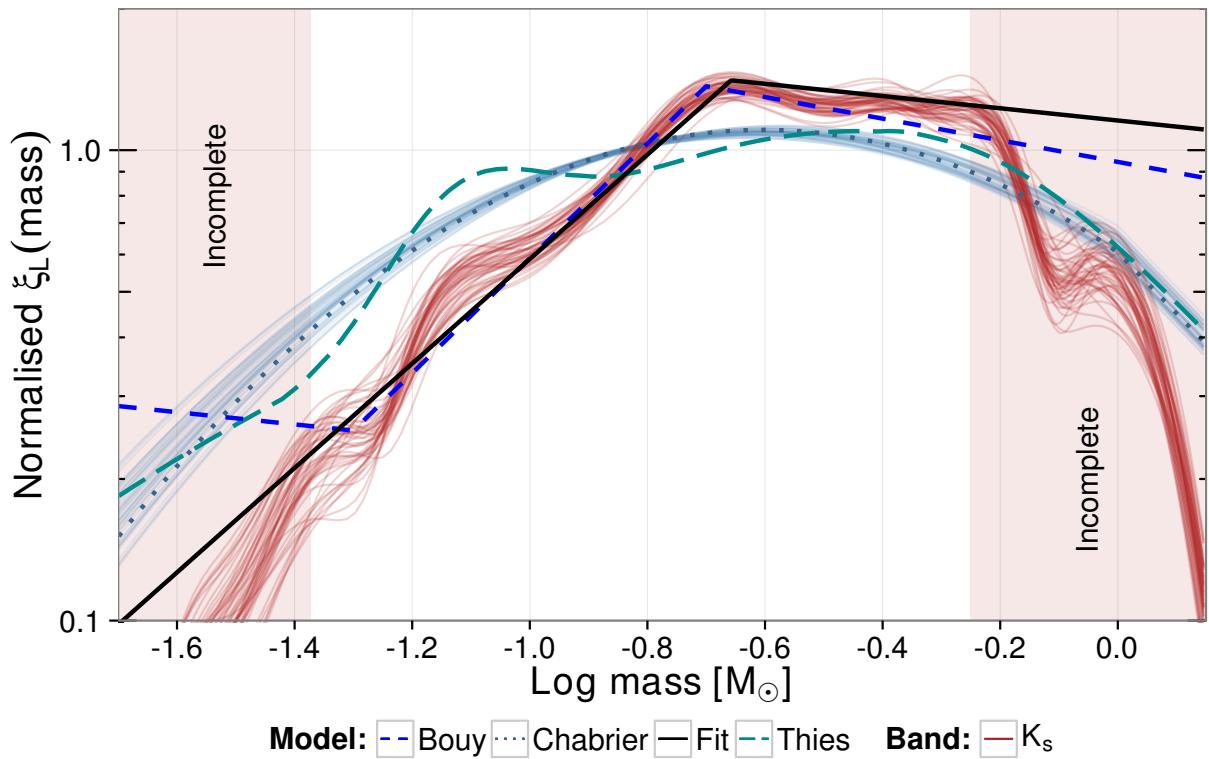


Figure 4.25: Normalised logarithmic PDSMD in  $K_s$  band. Also shown the IMFs of Chabrier (2005) (blue dotted line with uncertainties from Chabrier 2003a) and Thies & Kroupa (2007) (turquoise long-dashed line), and power-law models found here (black solid line, see text) and by Bouy et al. (2015) (blue dashed line).

and Hyades<sup>4</sup>. These PDSMDs correspond to those of Fig. 11 of Bouy et al. (2015). As mentioned by Bouy et al. (2015), the abundance of low-mass stars and brown dwarfs in the range  $0.03 - 0.1 M_{\odot}$  ( $-1.4 < \log M/M_{\odot} < -1$ ) seems to diminish with time. The relative increase of objects in the range  $-0.4 < \log M/M_{\odot} < -0.2$  is an effect of the normalisation<sup>5</sup>. This effect is consistent with the classical scenario in which low-mass stars and brown dwarfs are ejected as the cluster relaxes.

Since I lack the learned BHM for these two open clusters, the following comparison is made on a frequentist hypothesis testing approach, rather than on the proper Bayesian model selection scheme.

In this hypothesis test, the null hypothesis is that the Hyades and Trapezium PDSMDs came, each of them, from the same distribution than the Pleiades.

If we want to test the null hypothesis that two distributions come from the same parent distribution, Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD) tests are classical options, with the AD the most robust one. To perform these tests, we must compute certain measures from the two distributions. Then, given the measure, the test distribution returns the probability that the two distributions came from the same parent distribution. Finally, we reject the null hypothesis only if the previous probability is lower than certain probability threshold ( $\alpha$ ), which is usually 0.1, 0.05, or 0.01.

To perform the KS tests we must obtain the maximum distance between the CDFs of the two distributions. Then, using this distance and the KS distribution, we obtain the probability that the two distributions came from the same parent distribution.

However, the KS test can also be applied in a graphical way. Given the  $\alpha$  probability threshold, there is a  $d_{\alpha}$  distance for which the KS distribution returns a probability  $\alpha$ . For  $d < d_{\alpha}$   $p_{KS}(d) > \alpha$  and for  $d > d_{\alpha}$   $p_{KS}(d) < \alpha$ . Therefore, given the CDF of one of the distributions that we want to compare and a probability threshold  $\alpha$ , the region of distance  $d_{\alpha}$  around the CDF depicts the hypothesis test. If the CDF of the other distribution lies entirely within this region, then its maximum distance from the first CDF is less than  $d_{\alpha}$ . Therefore, its probability is greater than  $\alpha$  and the null hypothesis can not be rejected.

Figure 4.27 shows the cumulative distribution functions (CDFs) of the Trapezium, Pleiades (in  $K_s$  band) and Hyades PDSMDs. Also and for comparison, I show the CDFs resulting of Chabrier (2005) and Thies & Kroupa (2007) IMFs. The grey area around the Pleiades CDF depicts the graphical KS hypothesis test in which I choose  $\alpha = 0.01$ .

Furthermore, since the Kolmogorov-Smirnov test uses only the maximum distance between CDFs, I also applied the more robust Anderson-Darling test. It also rejects the null hypotheses (at  $p < 0.004$ ) that the Trapezium and Hyades PDSMDs and the Chabrier (2005) and Thies & Kroupa (2007) IMFs came from the same parent distribution as the Pleiades PDSMD.

The previous tests suggest that there is enough evidence for the observed differences

---

<sup>4</sup>Kindly provided by Hervé Bouy in a private communication

<sup>5</sup>The interesting alternative of open clusters gaining intermediate mass stars is yet to be explored.

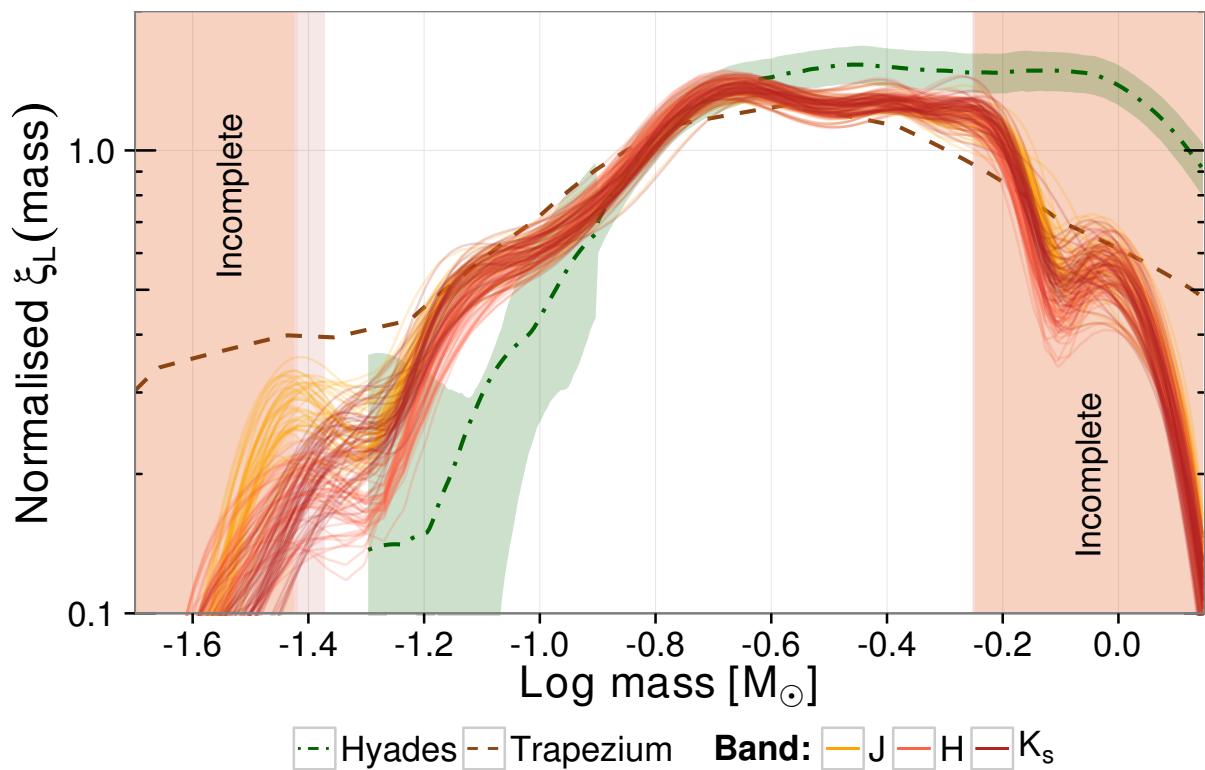


Figure 4.26: The PDSMDs of the Pleiades (derived here for  $J, H, K_s$  bands), Trapezium, and Hyades (both from Bouy et al. (2015)) clusters. They are normalised in the interval of completeness.

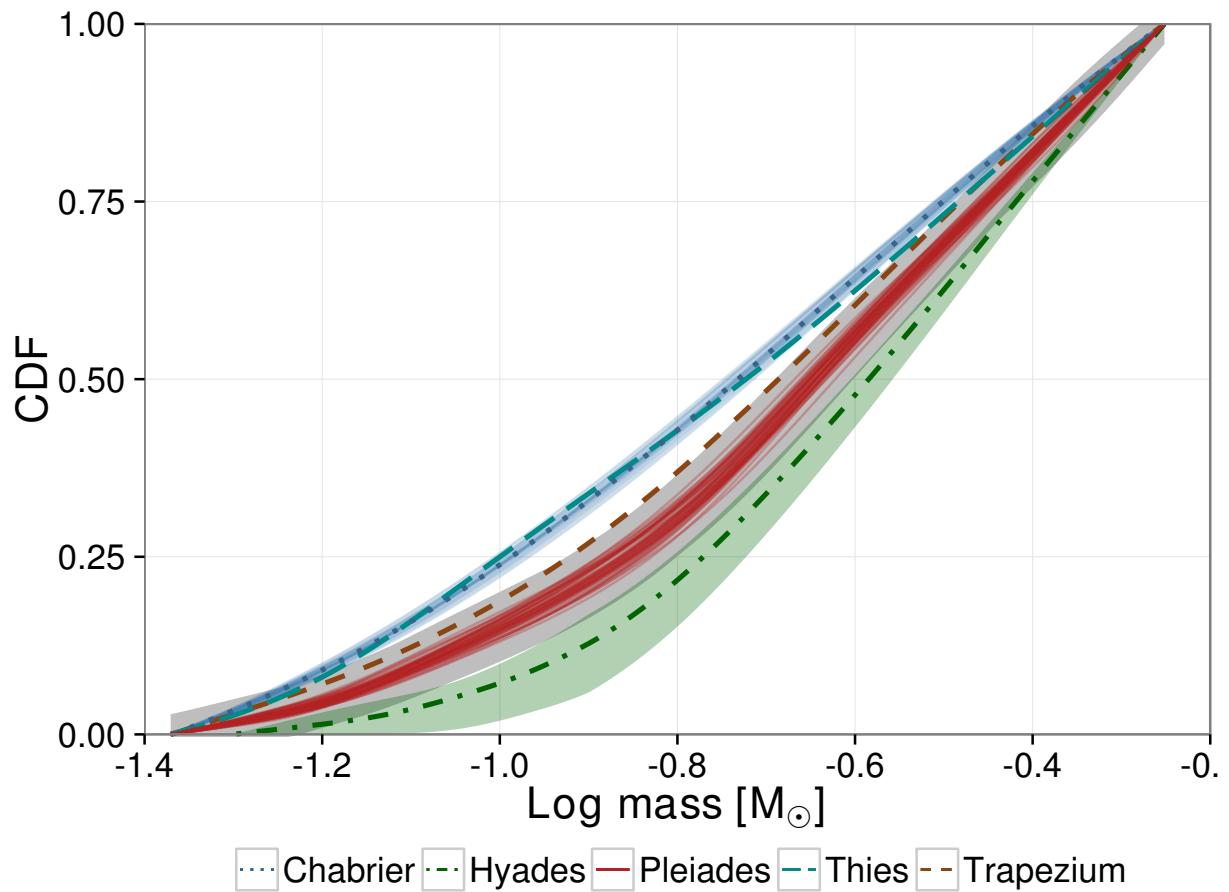


Figure 4.27: Cumulative distribution functions (CDF) of the PDSMDs from left panel and that of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#) system initial mass function (normalised also in the interval of completeness). The shown Pleiades CDF correspond to the  $K_s$  band. The grey area depicts the area in which the CDFs (both Trapezium and Hyades) should lie for the null hypothesis not to be rejected (at  $\alpha = 0.01$ ).

among the PDSMDs of these three clusters. Also, they suggest that IMFs of Chabrier (2005) and Thies & Kroupa (2007) are statistically different from the Pleiades PDSMD. These observed differences, as mentioned in the previous Section, may have its origin on dynamical effects associated with age and relaxation.

Nevertheless, to claim for reliable evidence supporting these differences, several issues must be solved. First, the uncertainties in the PDSMD must be properly established. Second, the luminosity distributions must include the uncertainties from the data not just from the poissonian counts. For this, the BHM of these cluster must be computed. Finally, the proper way to compare models is under Bayes' theorem.

# Chapter 5

## Conclusions and future work

In this work, I have created, tested, and validated a new Intelligent System (IS) that recovers the probability distribution functions (PDFs) of star clusters kinematic and photometric observables. In particular, the proper motions and apparent magnitudes. When the distance to the cluster is known, the PDFs of these observables can be transformed into transverse velocity and luminosity distributions. In addition, using the cluster age and a mass-luminosity relation the latter can be transformed into a mass distribution. A compilation of kinematic and mass distributions of Nearby Young Clusters (NYC) in diverse environments and spanning the ages of the early phases of cluster evolution will allow the astrophysical community to understand the effects that the environments and the internal dynamics have in the origin and evolution of stellar clusters. Since the majority of the stars form in clusters, understanding how these form and evolve will shed light into the knowledge of the past and future of our galaxy.

The IS presented in this work solves most of the issues found in current methodologies of cluster members determination and analysis. Furthermore, it steps ahead of the current paradigm of these methodologies, which is the classification of objects in cluster and field populations, and moves into a new one: *learning the kinematic and photometric distributions of the cluster populations*, both single and binary stars.

Thanks to its Bayesian framework, and to the Bayesian Hierarchical Model (BHM) methodology and the use of weakly informative priors specifically, this new IS solves the issues present in current classifiers from the literature and minimises their biases. In particular, it fulfils its objective by properly propagating the observable multivariate and heteroscedastic uncertainties, using the valuable information provided by objects with missing values in their observables, and eliminating the sampling bias associated with the use of only high membership probability objects.

To learn the posterior distributions of the parameters in the cluster model, the BHM, as we call this new IS, uses a combination of learning techniques, the Particle Swarm Optimiser of [Kennedy & Eberhart \(1995\)](#); [Clerc & Kennedy \(2002\)](#) and the *emcee* Markov Chain Monte Carlo method of [Foreman-Mackey et al. \(2013\)](#). Both are implemented in

High-Performance Computing environments to use large data sets ( $10^5$  to  $10^6$  objects) and multidimensional observable spaces. While the multidimensionality of the observable space (proper motions and photometry) add valuable information to constrain the cluster model and render it accurate, the large size of the data set increases its precision.

The performance of the BHM as a classifiers was analysed on synthetic data sets. The results of this analysis show that at an optimal probability threshold of  $p = 0.84$  the it has a True Positive Rate (recovery rate) of  $90.0 \pm 0.05\%$  and a contamination rate of  $4.3 \pm 0.2\%$ . The area under the receiving-operating characteristic curve (AUC) is 0.99 which indicates that it as an excellent classifier. The synthetic analysis reveals that the expected value of the contamination in the kinematic and photometric cluster distributions is  $5.8 \pm 0.2\%$ .

The BHM has been thoroughly tested and validated on the Pleiades DANCe Data Release 2 [Bouy et al. \(2015\)](#). Since the Pleiades is one of the most studied clusters in the history of astronomy, it offers the perfect case study for the benchmarking of an IS aiming at the analysis of NYC. In addition, the high precision and carefully designed observations of the DANCe project for the Pleiades cluster provide the necessary detailed information and sufficient statistics to benchmark the BHM and constrain its posterior distribution, in the high dimensionality (85) of its parametric space specifically.

The results of applying the BHM to the Pleiades DANCe DR2 data set yield the following astrophysical results.

- The by-product cluster membership probabilities, which are now delivered as full PDFs, show that (see Section 4.2):
  - There is an outstanding agreement of 99% with the classification process done by [Bouy et al. \(2015\)](#) on the approx. two million objects data set.
  - From the 1967 pleiads candidate members found in this work, 205 are new ones. It represents 10% of the cluster population.
  - From the candidate members of [Rebull et al. \(2016\)](#), 91% are classified as members. This figure is better than our estimated value of the true positive rate,  $90.0 \pm 0.05$ .
- The model selection analysis of the PSD performed on the candidate members of the BHM shows that (see Section 4.5):
  - The King’s profile ([King 1962](#)) is not just a good model due to its physical interpretability, but also has the largest Bayesian evidence when compared to classical profiles from the literature.
  - There is enough Bayesian evidence to support luminosity segregation of the Pleiades cluster.

- The Pleiades data set supports, with large Bayesian evidence, PSD models with biaxial symmetry indicating the non negligible ellipticity of the cluster ( $\epsilon = 0.13$ ).
- The kinematic and photometric distribution of the cluster populations show that (see Section 4.6 and 4.7):
  - The equal-mass binary (EMB) fraction is  $9.9 \pm 0.1\%$ .
  - The centroids of the cluster proper motions are  $\{16.30, -39.68\}$  mas · yr<sup>-1</sup> and  $\{15.71, -40.31\}$  mas · yr<sup>-1</sup> for single and EMB, respectively.
  - The derived luminosity distributions in the infrared bands ( $J$ ,  $H$ , and  $K_s$ ) in the completeness interval are in good agreement with the previous ones of [Bouy et al. \(2015\)](#).
- The derived present-day system mass distribution (using an age of 120 Myr and the BT-Settl isochrone model of [Allard et al. 2012](#)) shows (see Sections 4.8 and 4.9):
  - A general agreement with the IMFs of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#), in the intermediate mass range specifically.
  - The IMFs of [Chabrier \(2005\)](#) and [Thies & Kroupa \(2007\)](#) predict too many low-mass stars and BD.
  - A trend of depletion in low-mass stars and BD with cluster age, when comparing it with the Trapezium and Hyades PDSMD from the literature.

By minimising bias, propagating uncertainties, including correlations and doing a comprehensive modelling of the data particularities (missing values and heteroscedastic uncertainties) the BHM, developed, tested and validated in the present work, delivers the kinematic and photometric distributions of the cluster single and equal-mass binary populations. This new intelligent system represents the state of the art in the statistical analysis of Nearby Young Clusters.

## 5.1 Current issues and future improvements

Despite the fact that the BHM presented here solves the issues present in current methodologies from the literature, the following is a non-exhaustive list of important aspects to tackle in the near future.

- The BHM must be able to deal with:
  - Unresolved binaries of different mass ratios. Although the CSN has already been included in the BHM to deal with this issue, its application still demands fine tuning of the HPC elements (see Section 3.3.4).

- Proper motion and photometry from other surveys (e.g. *Pan-STARRS*, *Gaia*, *LSST*). Although so far untested, the BHM, thanks to its ability to cope with large amounts of missing values, it can deal with data sets from heterogeneous origins. For example, *Gaia* proper motion data (together with its correlated uncertainties) plus *Pan-STARRS* photometry, can be directly plugged into the BHM to improve the uncertainties of the kinematic and photometric cluster distributions.
- Other observables. As shown with the projected spatial distribution, other observables like radial velocities from the *Gaia-ESO* survey and parallaxes from *Gaia*, will give useful information to further refine the lists of candidate members.
- Extinction. The treatment of interstellar extinction is of paramount importance and is mandatory for the application of the BHM in younger and embedded clusters. Once done, it will allow the unravelling of kinematic and photometric distributions of embedded young clusters.
- Unresolved multiple systems. Although it is expected that these systems represent a small fraction of the cluster population (only 4% for the triple systems Duquennoy & Mayor 1991), they nevertheless have an important contribution to the mass distribution.
- Clusters with superimposed populations. The treatment of superimposed populations is also important to minimise contamination.
- The white dwarfs population. Although their numbers are negligible on NYC, their mass contribution nevertheless must be considered to properly constrain the PDSMD and the IMFs.
- The time needed to learn the BHM. Currently, the BHM takes four weeks to run in a 80 CPU computing cluster. Although the DANCe team has lately translated the CPU code into GPU code, the latter still needs improvement, in memory allocation and parallel distribution particularly.
- The choosing of prior distributions in other non-well studied NYC. In NYC where the prior information is missing, other kind of prior distributions must be used (e.g. objective or non-informative priors). Nevertheless, the BHM still provides the less subjective path to parametric inference.
- Choosing the appropriate number of GMM components. As with the previous point, where the prior information is not available, we can nevertheless use non-informative priors and large number of GMM components, so that those non-required components had small weight and covariance matrices. A model selection analysis must be implemented.

- Uncertainties to the mass-luminosity relation. Although non related to the BHM, the mass-luminosity relation is used univocally to transform a luminosity into mass. However, this knowledge is uncertain and its uncertainty must be incorporated into the derived mass distributions.

Once these issues will be solved, the improved BHM will render the inventory of cluster kinematic and luminosity distributions that will allow the DANCe team to fulfil the objective of understanding the effects that the environment (i.e. initial conditions) and the internal dynamics (e.g. evaporation, ejection, mass segregation) have on the clusters formation and evolution.

Finally, I list some future perspectives of the BHM.

- It will eventually become a free and open-source code for the use of the astronomical community.
- Although it has been created for the analysis of star clusters, it can also be applied to stellar associations.
- The empirical colour-magnitude relations that it learns can be used to constrain the theoretical evolutionary models once corrected by individual distances and extinctions.
- The lists of candidate members and equal-mass binaries that it delivers can be used to perform efficient searches and follow-ups.
- In the *Gaia* era, it will continue to provide not just accurate lists of candidate members, but also unbiased kinematic and photometric distributions with correctly propagated uncertainties for objects all along from the high mass domain down to the planetary mass regime.

# Bibliography

- Adams, J. D., Stauffer, J. R., Monet, D. G., Skrutskie, M. F., & Beichman, C. A. 2001, AJ, 121, 2053
- Adams, W. S. 1904, ApJ, 19
- Akeret, J., Seehars, S., Amara, A., Refregier, A., & Csillaghy, A. 2013, Astronomy and Computing, 2, 27
- Allard, F. 2014, in IAU Symposium, Vol. 299, Exploring the Formation and Evolution of Planetary Systems, ed. M. Booth, B. C. Matthews, & J. R. Graham, 271–272
- Allard, F., Homeier, D., & Freytag, B. 2013, Mem. Soc. Astron. Italiana, 84, 1053
- Allard, F., Homeier, D., Freytag, B., & Sharp, C. M. 2012, in EAS Publications Series, Vol. 57, EAS Publications Series, ed. C. Reylé, C. Charbonnel, & M. Schultheis, 3–43
- Alonso-Santiago, J., Negueruela, I., Marco, A., et al. 2017, MNRAS, 469, 1330
- Alsing, J., Heavens, A., & Jaffe, A. H. 2017, MNRAS, 466, 3272
- Anderes, E., Wandelt, B. D., & Lavaux, G. 2015, ApJ, 808, 152
- Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. I. 2003, Machine Learning, 50, 5
- Astraatmadja, T. L. & Bailer-Jones, C. A. L. 2016a, ApJ, 832, 137
- Astraatmadja, T. L. & Bailer-Jones, C. A. L. 2016b, ApJ, 833, 119
- Bailer-Jones, C. A. L. 2015, PASP, 127, 994
- Balaguer-Núñez, L., Galadí-Enríquez, D., & Jordi, C. 2007, A&A, 470, 585
- Ballesteros-Paredes, J., Hartmann, L. W., Pérez-Goytia, N., & Kuznetsova, A. 2015, MNRAS, 452, 566
- Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403

- Barnes, III, T. G., Moffett, T. J., Jefferys, W. H., & Forestell, A. D. 2004, in Astronomical Society of the Pacific Conference Series, Vol. 310, IAU Colloq. 193: Variable Stars in the Local Group, ed. D. W. Kurtz & K. R. Pollard, 95
- Barrado, D., Bouy, H., Bouvier, J., et al. 2016, *A&A*, 596, A113
- Bate, M. R. 2009a, *MNRAS*, 392, 590
- Bate, M. R. 2009b, *MNRAS*, 397, 232
- Bate, M. R. 2009c, *MNRAS*, 392, 1363
- Bate, M. R., Bonnell, I. A., & Bromm, V. 2003, *MNRAS*, 339, 577
- Berger, J. O., Bernardo, J. M., & Sun, D. 2009, *Ann. Statist.*, 37, 905
- Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)
- Bishop, C. 2006, Pattern Recognition and Machine Learning, Information Science and Statistics (Springer)
- Blaauw, A. 1964, in IAU Symposium, Vol. 20, The Galaxy and the Magellanic Clouds, ed. F. J. Kerr, 50
- Blackwell, T. & Bentley, P. 2002, in Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computations, ed. W. Langdon, E. Cantu-Paz, K. Mathias, R. Roy, & D. Davis (San Francisco, CA, USA: Morgan kaufmann Publishers Inc.), 19–26
- Boss, L. 1910, Preliminary General Catalogue of 6188 stars for the epoch 1900
- Bouy, H., Bertin, E., Moraux, E., et al. 2013, *A&A*, 554, A101
- Bouy, H., Bertin, E., Sarro, L. M., et al. 2015, *A&A*, 577, A148
- Bovy, J., Hogg, D. W., & Roweis, S. T. 2009, *ApJ*, 700, 1794
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Brooks, S., Gelman, A., Jones, G., & Meng, X. 2011, Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC Handbooks of Modern Statistical Methods (CRC Press)
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, 564, A125
- Byun, Y.-I., Grillmair, C. J., Faber, S. M., et al. 1996, *AJ*, 111, 1889
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245

- Carpenter, B., Gelman, A., Hoffman, M., et al. 2017, Journal of Statistical Software, Articles, 76, 1
- Carpenter, J. M. 2000, AJ, 120, 3139
- Cartwright, A. & Whitworth, A. P. 2012, MNRAS, 423, 1018
- Chabrier, G. 2003a, PASP, 115, 763
- Chabrier, G. 2003b, ApJ, 586, L133
- Chabrier, G. 2005, in Astrophysics and Space Science Library, Vol. 327, The Initial Mass Function 50 Years Later, ed. E. Corbelli, F. Palla, & H. Zinnecker, 41
- Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., & Dorie, V. 2015, Journal of Educational and Behavioral Statistics, 40, 136
- Clerc, M. & Kennedy, J. 2002, Trans. Evol. Comp, 6, 58
- Converse, J. M. & Stahler, S. W. 2008, ApJ, 678, 431
- Converse, J. M. & Stahler, S. W. 2010, MNRAS, 405, 666
- Dahn, C. C., Liebert, J., & Harrington, R. S. 1986, AJ, 91, 621
- De Boor, C. 1978, A practical guide to splines / Carl de Boor (Springer-Verlag New York), xxiv, 392 p. :
- de Graeve, E. 1979, Publications of the Vatican Observatory V.1:16, P. 1, 1979, 1, 1
- de Marchi, G. & Paresce, F. 2001, in Astronomische Gesellschaft Meeting Abstracts, Vol. 18, Astronomische Gesellschaft Meeting Abstracts, ed. E. R. Schielicke
- de Zeeuw, T. 1985, MNRAS, 216, 273
- Deacon, N. R. & Hambly, N. C. 2004, A&A, 416, 125
- Delfosse, X., Forveille, T., Ségransan, D., et al. 2000, A&A, 364, 217
- Demory, B.-O. 2014, ApJ, 789, L20
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B, 39, 1
- Diard, J. & Bessiere, P. 2008, Probabilistic reasoning and decision making in sensory-motor systems, 153
- Dirac, P. A. M. 1942, Proceedings of the Royal Society of London Series A, 180, 1

- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, Physics Letters B, 195, 216
- Duquennoy, A. & Mayor, M. 1991, A&A, 248, 485
- Elson, R. A. W., Fall, S. M., & Freeman, K. C. 1987, ApJ, 323, 54
- Faherty, J. K., Burgasser, A. J., Walter, F. M., et al. 2012, ApJ, 752, 56
- Fan, Y. & Sisson, S. 2011, Handbook of Markov Chain Monte Carlo, 67
- Feeney, S. M., Johnson, M. C., McEwen, J. D., Mortlock, D. J., & Peiris, H. V. 2013, Phys. Rev. D, 88, 043012
- Fei-Fei, L. & Perona, P. 2005, in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, Ieee, 524–531
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601
- Flegal, J. M., Hughes, J., & Vats, D. 2016, mcmcse: Monte Carlo Standard Errors for MCMC, Riverside, CA and Minneapolis, MN, r package version 1.2-1
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the Astronomical Society of the Pacific, 125, 306
- Franklin, J. 2002, The Science of Conjecture: Evidence and Probability Before Pascal, A Johns Hopkins paperback (Johns Hopkins University Press)
- Gagné, J., Lafrenière, D., Doyon, R., Malo, L., & Artigau, É. 2014, ApJ, 783, 121
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016a, A&A, 595, A1
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016b, A&A, 595, A1
- Gaia Collaboration, van Leeuwen, F., Vallenari, A., et al. 2017, A&A, 601, A19
- Galli, P. A. B., Moraux, E., Bouy, H., et al. 2017, A&A, 598, A48
- Gatewood, G., de Jonge, J. K., & Han, I. 2000, ApJ, 533, 938
- Gelman, A. 2006, Bayesian Analysis, 1, 515
- Gelman, A. 2012, CHANCE, 25, 52
- Gelman, A., Carlin, J., Stern, H., et al. 2013, Bayesian Data Analysis, third edition edn. (CRC)

- Gelman, A. & Hill, J. 2007, Data Analysis Using Regression and Multilevel/Hierarchical Models, Analytical Methods for Social Research (Cambridge University Press)
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. 2008, Ann. Appl. Stat., 2, 1360
- Geman, S. & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721
- Gong, L. & Flegal, J. M. 2016, Journal of Computational and Graphical Statistics, 25, 684
- González-Farias, G., Dominguez-Molina, A., & Gupta, A. K. 2004, Journal of Statistical Planning and Inference, 126, 521
- Good, I. J. 1980, Trabajos de Estadistica Y de Investigacion Operativa, 31, 489
- Goodman, J. & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
- Gupta, A. K., González-Farias, G., & Dominguez-Molina, J. 2004, Journal of Multivariate Analysis, 89, 181
- Guthrie, B. N. G. 1987, QJRAS, 28, 289
- Hambly, N. C. & Jameson, R. F. 1991, MNRAS, 249, 137
- Hastings, W. K. 1970, Biometrika, 57, 97
- Henry, T. J. & McCarthy, Jr., D. W. 1990, ApJ, 350, 334
- Hernquist, L. 1990, ApJ, 356, 359
- Hertzsprung, E. 1947, Annalen van de Sterrewacht te Leiden, 19, A1
- Hillenbrand, L. A., Bauermeister, A., & White, R. J. 2008, in 14th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. G. van Belle, Vol. 384, 200
- Hodgkin, S. T. & Jameson, R. F. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 198, Stellar Clusters and Associations: Convection, Rotation, and Dynamos, ed. R. Pallavicini, G. Micela, & S. Sciortino, 59
- Hogg, D. W., Bovy, J., & Lang, D. 2010a, ArXiv e-prints
- Hogg, D. W., Myers, A. D., & Bovy, J. 2010b, ApJ, 725, 2166
- Holl, B., Hobbs, D., & Lindegren, L. 2010, in IAU Symposium, Vol. 261, Relativity in Fundamental Astronomy: Dynamics, Reference Frames, and Data Analysis, ed. S. A. Klioner, P. K. Seidelmann, & M. H. Soffel, 320–324
- Hong, D., Balzano, L., & Fessler, J. A. 2016, Towards a theoretical analysis of PCA for heteroscedastic data, arxiv 1610.03595

- Huang, A. & Wand, M. P. 2013, Bayesian Analysis, 8, 439
- Jaffe, W. 1983, MNRAS, 202, 995
- Jameson, R. F., Dobbie, P. D., Hodgkin, S. T., & Pinfield, D. J. 2002, MNRAS, 335, 853
- Jappsen, A.-K., Klessen, R. S., Larson, R. B., Li, Y., & Mac Low, M.-M. 2005, A&A, 435, 611
- Jaynes, E. 2003, Probability Theory: The Logic of Science (Cambridge University Press)
- JCGM. 2008, JCGM 100: Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement, Tech. rep., Joint Committee for Guides in Metrology
- Jefferys, T. R., Jefferys, W. H., Barnes, III, T. G., & Dambis, A. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 371, Statistical Challenges in Modern Astronomy IV, ed. G. J. Babu & E. D. Feigelson, 433
- Jeffreys, H. 1961, Theory of Probability, 3rd edn. (Oxford, England: Oxford)
- Jiang, F. & van den Bosch, F. C. 2016, MNRAS, 458, 2848
- Johnson, H. L. & Mitchell, R. I. 1958, ApJ, 128, 31
- Jones, B. F. 1970, AJ, 75, 563
- Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
- Kass, R. E. & Raftery, A. E. 1995, Journal of the American Statistical Association, 90, 773
- Kennedy, J. & Eberhart, R. 1995, in Neural Networks, 1995. Proceedings., IEEE International Conference on, Vol. 4, 1942–1948 vol.4
- Kennicutt, Jr., R. C. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 142, The Stellar Initial Mass Function (38th Herstmonceux Conference), ed. G. Gilmore & D. Howell, 1
- King, I. 1962, AJ, 67, 471
- Koller, D. & Friedman, N. 2009, Probabilistic Graphical Models: Principles and Techniques, Adaptive computation and machine learning (MIT Press)
- Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, AJ, 146, 134
- Kozhurina-Platais, V., Girard, T. M., Platais, I., et al. 1995, AJ, 109, 672
- Krone-Martins, A. & Moitinho, A. 2014, A&A, 561, A57
- Kroupa, P. 2001, MNRAS, 322, 231

- Kroupa, P. 2002, Science, 295, 82
- Kroupa, P., Gilmore, G., & Tout, C. A. 1991, MNRAS, 251, 293
- Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. 2013, The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations, ed. T. D. Oswalt & G. Gilmore, 115
- Kuznetsova, A., Hartmann, L., & Ballesteros-Paredes, J. 2015, ApJ, 815, 27
- Lada, C. J. & Lada, E. A. 2003, ARA&A, 41, 57
- Lauer, T. R., Ajhar, E. A., Byun, Y.-I., et al. 1995, AJ, 110, 2622
- Lee, S.-W. & Sung, H. 1995, Journal of Korean Astronomical Society, 28, 45
- Limber, D. N. 1960, ApJ, 131, 168
- Limber, D. N. 1961, ApJ, 134, 537
- Limber, I. D. N. 1962, ApJ, 135, 16
- Lin, T., Lee, J., & Ho, H. 2006, WOS:000236696400015
- Lindstrom, M. 1999, Journal of Computational and Graphical Statistics, 8, 333
- Liu, T., Janes, K. A., & Bania, T. M. 1991, ApJ, 377, 141
- Lodieu, N., Deacon, N. R., & Hambly, N. C. 2012, MNRAS, 422, 1495
- Lodieu, N., Dobbie, P. D., Deacon, N. R., et al. 2007, MNRAS, 380, 712
- Loktin, A. V. 2006, Astronomy Reports, 50, 714
- Lutz, T. E. & Kelker, D. H. 1973, PASP, 85, 573
- Luyten, W. J. 1939, Publications of the Astronomical Observatory University of Minnesota, 7, 121
- Luyten, W. J. 1941, Annals of the New York Academy of Sciences, 42, 201
- Maíz-Apellániz, J. 2003, in HST Calibration Workshop : Hubble after the Installation of the ACS and the NICMOS Cooling System, ed. S. Arribas, A. Koekemoer, & B. Whitmore, 346
- Malo, L., Doyon, R., Lafrenière, D., et al. 2013, ApJ, 762, 88
- March, M. C., Karpenka, N. V., Feroz, F., & Hobson, M. P. 2014, MNRAS, 437, 3298
- Martin, E. L., Rebolo, R., & Zapatero-Osorio, M. R. 1996, ApJ, 469, 706

- Maschberger, T. 2013, MNRAS, 429, 1725
- McMichael, D. W. 1996, in Proc. Fourth International Symposium on Signal Processing and its Applications (ISSPA, 377–378
- Melis, C., Reid, M. J., Mioduszewski, A. J., Stauffer, J. R., & Bower, G. C. 2014, Science, 345, 1029
- Mermilliod, J.-C. 1979, Bulletin d'Information du Centre de Donnees Stellaires, 16, 2
- Mermilliod, J.-C., Bratschi, P., & Mayor, M. 1997, A&A, 320, 74
- Mermilliod, J.-C., Mayor, M., & Udry, S. 2009, A&A, 498, 949
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087
- Miller, G. E. & Scalo, J. M. 1979, ApJS, 41, 513
- Moore, B., Quinn, T., Governato, F., Stadel, J., & Lake, G. 1999, MNRAS, 310, 1147
- Moraux, E., Bouvier, J., & Stauffer, J. R. 2001, A&A, 367, 211
- Moraux, E., Bouvier, J., Stauffer, J. R., & Cuillandre, J.-C. 2003, A&A, 400, 891
- Moraux, E., Kroupa, P., & Bouvier, J. 2004, A&A, 426, 75
- Morris, C. N. 1983, Journal of the American Statistical Association, 78, 47
- Mulders, G. D., Pascucci, I., & Apai, D. 2015, ApJ, 814, 130
- Myeong, G. C., Jerjen, H., Mackey, D., & Da Costa, G. S. 2017, ApJ, 840, L25
- Narayanan, D. & Davé, R. 2012, MNRAS, 423, 3601
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
- Neal, R. M. 1996, Bayesian Learning for Neural Networks (Secaucus, NJ, USA: Springer-Verlag New York, Inc.)
- Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, Protostars and Planets VI, 53
- Olivares, J., Sánchez, L. J., Ruelas-Mayorga, A., et al. 2013, AJ, 146, 106
- Olivares, J., Sarro, L., Moraux, E., Berihuete, A., & Bouy, H. 2017, A&A, submitted
- Panwar, N., Samal, M. R., Pandey, A. K., et al. 2017, MNRAS, 468, 2684
- Patel, P. K., Sharma, V., & Gupta, K. 2013, International Journal of Computer Applications, 73

- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, 323, L49
- Pinfield, D. J., Jameson, R. F., & Hodgkin, S. T. 1998, *MNRAS*, 299, 955
- Plummer, H. C. 1911, *MNRAS*, 71, 460
- Porras, A., Christopher, M., Allen, L., et al. 2003, *AJ*, 126, 1916
- Press, W. H. 1997, in *Unsolved Problems in Astrophysics*, ed. J. N. Bahcall & J. P. Ostriker, 49–60
- Pritchard, C. 1884, *MNRAS*, 44, 355
- Raboud, D. & Mermilliod, J.-C. 1998, *A&A*, 329, 101
- Rebull, L. M., Stauffer, J. R., Bouvier, J., et al. 2016, *AJ*, 152, 113
- Reid, I. N., Cruz, K. L., Allen, P., et al. 2004, *AJ*, 128, 463
- Reid, I. N., Gizis, J. E., & Hawley, S. L. 2002, *AJ*, 124, 2721
- Riedel, A. R., Blunt, S. C., Lambrides, E. L., et al. 2017, *AJ*, 153, 95
- Roberts, G. O. & Rosenthal, J. S. 2004, *Probability Surveys*, 1, 20
- Robotham, A. S. G., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. P. 2017, *MNRAS*, 466, 1513
- Roswick, J. M., Mermilliod, J.-C., & Mayor, M. 1992, *A&A*, 255, 130
- Sackett, P. D. & Sparke, L. S. 1990, *ApJ*, 361, 408
- Sale, S. E. 2012, *MNRAS*, 427, 2119
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Sampedro, L. & Alfaro, E. J. 2016, *MNRAS*, 457, 3949
- Sanders, W. L. 1971, *A&A*, 14, 226
- Sarro, L. M., Berihuete, A., Carrión, C., et al. 2013, *A&A*, 550, A44
- Sarro, L. M., Bouy, H., Berihuete, A., et al. 2014, *Astronomy & Astrophysics*, 14
- Scalo, J. M. 1986, *Fund. Cosmic Phys.*, 11, 1
- Schwarz, G. 1978, *Ann. Statist.*, 6, 461
- Shariff, H., Dhawan, S., Jiao, X., et al. 2016, *MNRAS*, 463, 4311

- Shkedy, Z., Decin, L., Molenberghs, G., & Aerts, C. 2007, MNRAS, 377, 120
- Simon, M. 1997, The Astrophysical Journal Letters, 482, L81
- Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405
- Skilling, J. 2006, Bayesian Anal., 1, 833
- Smart, W. M. & Green, E. b. R. M. 1977, Textbook on Spherical Astronomy
- Smith, B. & Struve, O. 1944, ApJ, 100, 360
- Soderblom, D. R., Nelan, E., Benedict, G. F., et al. 2005, AJ, 129, 1616
- Sohn, S. T., Majewski, S. R., Muñoz, R. R., et al. 2007, ApJ, 663, 960
- Spiriti, S., Eubank, R., Smith, P. W., & Young, D. 2013, Journal of Statistical Computation and Simulation, 83, 1020
- Stauffer, J. R., Hartmann, L. W., Fazio, G. G., et al. 2007, ApJS, 172, 663
- Stauffer, J. R., Liebert, J., Giampapa, M., et al. 1994, AJ, 108, 160
- Stauffer, J. R., Schultz, G., & Kirkpatrick, J. D. 1998, ApJ, 499, L199
- Takeda, Y., Hashimoto, O., & Honda, S. 2017, PASJ, 69, 1
- Tapiador, D., Berihuete, A., Sarro, L. M., Julbe, F., & Huedo, E. 2017, Astronomy and Computing, 19, 1
- Terndrup, D. M., Stauffer, J. R., Pinsonneault, M. H., et al. 2000, AJ, 119, 1303
- Thies, I. & Kroupa, P. 2007, ApJ, 671, 767
- Thies, I. & Kroupa, P. 2008, MNRAS, 390, 1200
- Titus, J. 1938, AJ, 47, 25
- Trumpler, R. J. 1921, Lick Observatory Bulletin, 10, 110
- van de Ven, G., van den Bosch, R. C. E., Verolme, E. K., & de Zeeuw, P. T. 2006, A&A, 445, 513
- van Leeuwen, F. 1980, in IAU Symposium, Vol. 85, Star Clusters, ed. J. E. Hesser, 157–162
- van Leeuwen, F. 1999, A&A, 341, L71
- van Leeuwen, F. 2009, A&A, 497, 209

- van Rhijn, P. J. 1925, Publications of the Kapteyn Astronomical Laboratory Groningen, 38
- van Rhijn, P. J. 1936, Publications of the Kapteyn Astronomical Laboratory Groningen, 47, 1
- Vasilevskis, S., Klemola, A., & Preston, G. 1958, AJ, 63, 387
- Wells, A. E. 1924, PhD thesis, THE UNIVERSITY OF MICHIGAN.
- Wheeler, C. 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts #213, 202
- Wolfgang, A. & Lopez, E. 2015, ApJ, 806, 183
- Woolley, R. V. D. R. 1956, MNRAS, 116, 296
- Zhao, H. 1997, MNRAS, 287, 525